



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number 121207

TO: Manjunath N Rao
Location: rem/3b813c70
Art Unit: 1652
Thursday, May 13, 2004

Case Serial Number: 09/930440

From: Toby Port
Location: Biotech-Chem Library
Remsen 1A59
Phone: 571-272-2523

toby.port@uspto.gov

Search Notes

Dear Examiner Rao,

Here are the results of your search.
Please feel free to contact me if you have any questions.

Toby Port



STIC SEARCH RESULTS FEEDBACK FORM

Biotech-Chem Library

Questions about the scope or the results of the search? Contact *the searcher or contact:*

Mary Hale, Information Branch Supervisor
571-272-2507 Remsen E01 D86

Voluntary Results Feedback Form

➤ I am an examiner in Workgroup: Example: 1610

➤ Relevant prior art **found**, search results used as follows:

- ☐ 102 rejection
- ☐ 103 rejection
- ☐ Cited as being of interest.
- ☐ Helped examiner better understand the invention.
- ☐ Helped examiner better understand the state of the art in their technology.

Types of relevant prior art found:

- ☐ Foreign Patent(s)
- ☐ Non-Patent Literature
(journal articles, conference proceedings, new product announcements etc.)

➤ Relevant prior art **not found**:

- ☐ Results verified the lack of relevant prior art (helped determine patentability).
- ☐ Results were not useful in determining patentability or understanding the invention.

Comments:

Drop off or send completed forms to STIC/Biotech-Chem Library Remsen Bldg.



Db 61 ATCATTCGCCAGATCGCCAGAACACACAGGCGAGCTGAGCTAGAGCCAAAGCGCATGATC 120
Qy 121 CGCATGCGCCAGAGAGTGGGGCTGATTTGTGCCAACTTCCAGAAAGAGTGAATTAATTC 180
Db 121 CGCATGCGCCAGAGAGTGGGGCTGATTTGTGCCAACTTCCAGAAAGAGTGAATTAATTC 180
Qy 181 AAGTTAATCGAAAGCTTGGAGAGGCCATACACTCGAACATTTCTGGGGGAAGAGC 240
Db 181 AAGTTAATCGAAAGCTTGGAGAGGCCATACACTCGAACATTTCTGGGGGAAGAGC 240
Qy 241 TACGGGAGACCAAAAGACATCTGAGATTCAAGCCATGACCATGACGGAGCTGAGAGC 300
Db 241 TACGGGAGACCAAAAGACATCTGAGATTCAAGCCATGACCATGACGGAGCTGAGAGC 300
Qy 301 TACGGCGAGAGAGTGGAGATCTTCTTCACTGCTCGCATGAGATGAGATGCGAGTTGA 360
Db 301 TACGGCGAGAGAGTGGAGATCTTCTTCACTGCTCGCATGAGATGAGATGCGAGTTGA 360
Qy 361 TTCTTGATGAACTGAAATGTTCCATTTTCAAGTTGAACTGAGAGACCTAATTAATTT 420
Db 361 TTCTTGATGAACTGAAATGTTCCATTTTCAAGTTGAACTGAGAGACCTAATTAATTT 420
Qy 421 CCTTATCTGAAAAGACACCAAAAAAGGTGGCCCAATGATGATCTCCAGTGGAGTGAAG 480
Db 421 CCTTATCTGAAAAGACACCAAAAAAGGTGGCCCAATGATGATCTCCAGTGGAGTGAAG 480
Qy 481 TCATGAGACACATGAGAGAGGTTATATCATGATCTGAGAGAGCCCTCAACCCCACTTCTGC 540
Db 481 TCATGAGACACATGAGAGAGGTTATATCATGATCTGAGAGAGCCCTCAACCCCACTTCTGC 540
Qy 541 TTCTTGATGAACTGAAATGTTCCATTTTCAAGTTGAACTGAGAGACCTAATTAATTT 600
Db 541 TTCTTGATGAACTGAAATGTTCCATTTTCAAGTTGAACTGAGAGACCTAATTAATTT 600
Qy 601 TCGGAATATCAGAAAGCTTCTTCCATTTCCATGAGGATCTTCTGAGCATGAAACAGGC 660
Db 601 TCGGAATATCAGAAAGCTTCTTCCATTTCCATGAGGATCTTCTGAGCATGAAACAGGC 660
Qy 661 ATAGGAGATCTGTGGCCGAGTGGCTCTGGGGGCGCAAGGTGTGGAAGTCACTACT 720
Db 661 ATAGGAGATCTGTGGCCGAGTGGCTCTGGGGGCGCAAGGTGTGGAAGTCACTACT 720
Qy 721 TGGACAAAGCTGGAAGGGAGTGAACACTCGGCTCGCTGAGGCTGAGAACTGAGC 780
Db 721 TGGACAAAGCTGGAAGGGAGTGAACACTCGGCTCGCTGAGGCTGAGAACTGAGC 780
Qy 781 GAGCTGGTGGCTCAGTGGCTCTTGTGAAGCTGGCTTCCCAACCAAGAGCTG 840
Db 781 GAGCTGGTGGCTCAGTGGCTCTTGTGAAGCTGGCTTCCCAACCAAGAGCTG 840
Qy 841 CTGCGCTGAGATGAGGCTGAGATGAGAGTGGGCAAGTCTGTGCTGAGCAAGTGA 900
Db 841 CTGCGCTGAGATGAGGCTGAGATGAGAGTGGGCAAGTCTGTGCTGAGCAAGTGA 900
Qy 901 ATTCCGGAAGGACCAATTTCAATGACATGCTCAACCGTGAAGTGGTGAAGCCCAA 960
Db 901 ATTCCGGAAGGACCAATTTCAATGACATGCTCAACCGTGAAGTGGTGAAGCCCAA 960
Qy 961 GCTTATCTGCTGAGAGACCTTTTAATCTAGTGGGGAAGAGTCTGCTGAGTGA 1020
Db 961 GCTTATCTGCTGAGAGACCTTTTAATCTAGTGGGGAAGAGTCTGCTGAGTGA 1020
Qy 1021 GAGGATGACACATCTGAGAGATTTGATGATTAATCATGCAAAAAATCAAGTCTTA 1080
Db 1021 GAGGATGACACATCTGAGAGATTTGATGATTAATCATGCAAAAAATCAAGTCTTA 1080

RESULT 2
US-09-930-440B-5
Sequence 5, Application US/09930440B
Patent No. US20020142386A1
GENERAL INFORMATION:
APPLICANT: Betendagh et al.

FILE REFERENCE: PF09P2
CURRENT APPLICATION NUMBER: US/09/930,440B
CURRENT FILING DATE: 2001-08-16
PRIOR APPLICATION NUMBER: 60/227,579
PRIOR FILING DATE: 2000-08-25
PRIOR APPLICATION NUMBER: 09/516,793
PRIOR FILING DATE: 2000-03-01
PRIOR APPLICATION NUMBER: 60/169,624
PRIOR FILING DATE: 1999-12-08
PRIOR APPLICATION NUMBER: 60/122,582
PRIOR FILING DATE: 1999-03-02
NUMBER OF SEQ ID NOS: 8
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 5
LENGTH: 1080
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (1) ..(1080)
US-09-930-440B-5

Query Match 100.0%; Score 1080; DB 9; Length 1080;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1080; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGCGCTGAGAGCTGAGAGCTGTGTCGCGGCGCTGAGTGGGCGGCAACACCGTGTCTC 60
Db 1 ATGCGCTGAGAGCTGAGAGCTGTGTCGCGGCGCTGAGTGGGCGGCAACACCGTGTCTC 60
Qy 61 ATCATGCGGAGATCGGCGAGAACCAACAGGCGGACCTGAGAGTGAAGCGCATGATC 120
Db 61 ATCATGCGGAGATCGGCGAGAACCAACAGGCGGACCTGAGAGTGAAGCGCATGATC 120
Qy 121 CGCATGCGCCAGAGAGTGGGGCTGATTTGTGCCAACTTCCAGAAAGAGTGAATTAATTC 180
Db 121 CGCATGCGCCAGAGAGTGGGGCTGATTTGTGCCAACTTCCAGAAAGAGTGAATTAATTC 180
Qy 181 AAGTTAATCGAAAGCTTGGAGAGGCCATACACTCGAACATTTCTGGGGGAAGAGC 240
Db 181 AAGTTAATCGAAAGCTTGGAGAGGCCATACACTCGAACATTTCTGGGGGAAGAGC 240
Qy 241 TACGGGAGACCAAAAGACATCTGAGATTCAAGCCATGACCATGACGGAGCTGAGAGC 300
Db 241 TACGGGAGACCAAAAGACATCTGAGATTCAAGCCATGACCATGACGGAGCTGAGAGC 300
Qy 301 TACGGCGAGAGAGTGGAGATCTTCTTCACTGCTCGCATGAGATGAGATGCGAGTTGA 360
Db 301 TACGGCGAGAGAGTGGAGATCTTCTTCACTGCTCGCATGAGATGAGATGCGAGTTGA 360
Qy 361 TTCTTGATGAACTGAAATGTTCCATTTTCAAGTTGAACTGAGAGACCTAATTAATTT 420
Db 361 TTCTTGATGAACTGAAATGTTCCATTTTCAAGTTGAACTGAGAGACCTAATTAATTT 420
Qy 421 CCTTATCTGAAAAGACACCAAAAAAGGTGGCCCAATGATGATCTCCAGTGGAGTGAAG 480
Db 421 CCTTATCTGAAAAGACACCAAAAAAGGTGGCCCAATGATGATCTCCAGTGGAGTGAAG 480
Qy 481 TCATGAGACACATGAGAGAGGTTATATCATGATCTGAGAGAGCCCTCAACCCCACTTCTGC 540
Db 481 TCATGAGACACATGAGAGAGGTTATATCATGATCTGAGAGAGCCCTCAACCCCACTTCTGC 540
Qy 541 TTCTTGATGAACTGAAATGTTCCATTTTCAAGTTGAACTGAGAGACCTAATTAATTT 600
Db 541 TTCTTGATGAACTGAAATGTTCCATTTTCAAGTTGAACTGAGAGACCTAATTAATTT 600
Qy 601 TCGGAATATCAGAAAGCTTCTTCCATTTCCATGAGGATCTTCTGAGCATGAAACAGGC 660
Db 601 TCGGAATATCAGAAAGCTTCTTCCATTTCCATGAGGATCTTCTGAGCATGAAACAGGC 660
Qy 661 ATAGGAGATCTGTGGCCGAGTGGCTCTGGGGGCGCAAGGTGTGGAAGTCACTACT 720

Thu May 13 11:53:07 2004

us-09-930-440b-5.rnpb

Page 3

Db 661 ATAGCATATCTGTGCGCCGAGTGGCTCTGCGGGCCAAAGGTGTGGAACCTCACTAACT 720
Qy 721 TTGACACAGACCTGGAAGGGGAGTGAACCACTCGGCTCGCTGAGAGCTGGAAGTGGCC 780
Db 721 TTGACACAGACCTGGAAGGGGAGTGAACCACTCGGCTCGCTGAGAGCTGGAAGTGGCC 780
Qy 781 GAGCTGTGTGCGGTCAAGTGCCTTGTGTGAGCGTGCCTTGGGCTTCCCAACCAAGCACTG 840
Db 781 GAGCTGTGTGCGGTCAAGTGCCTTGTGTGAGCGTGCCTTGGGCTTCCCAACCAAGCACTG 840
Qy 841 CTGCTCTGTGAGATGCTGCTGATGAGAGCTGGGCAAGTCTGTGTGCGCAAGTGA 900
Db 841 CTGCTCTGTGAGATGCTGCTGATGAGAGCTGGGCAAGTCTGTGTGCGCAAGTGA 900
Qy 901 ATTCCGGAAGGACCATTTTAAACATGAGATGCTCAAGTGAAGTGGTGGAGCCCAAA 960
Db 901 ATTCCGGAAGGACCATTTTAAACATGAGATGCTCAAGTGAAGTGGTGGAGCCCAAA 960
Qy 961 GCTATCTCTCTGAAGACATCTTTAATCTAGTGGGCAAGAGTCTGTGCTGCTGAA 1020
Db 961 GCTATCTCTCTGAAGACATCTTTAATCTAGTGGGCAAGAGTCTGTGCTGCTGAA 1020
Qy 1021 GAGGATGACACCATGATGAGAGATGCTGATGATGATGATGATGATGATGATGATGAT 1080
Db 1021 GAGGATGACACCATGATGAGAGATGCTGATGATGATGATGATGATGATGATGATGAT 1080

RESULT 3

US-10-264-237-1192
Sequence 1192, Application US/10264237
Publication No. US20040008491A1
GENERAL INFORMATION:
APPLICANT: Bize et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PA3131P1
CURRENT APPLICATION NUMBER: US/10/264,237
CURRENT FILING DATE: 2002-10-04
PRIORITY APPLICATION NUMBER: PCT/US01/16450
PRIORITY FILING DATE: 2001-05-18
PRIOR APPLICATION NUMBER: US 60/205,515
PRIOR FILING DATE: 2000-05-19
NUMBER OF SEQ ID NOS: 2876
SOFTWARE: PatentIn Ver. 3.1
SEQ ID NO 1192
LENGTH: 1268
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (5)..(5)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (17)..(17)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (23)..(23)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (31)..(31)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1224)..(1227)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc_feature

LOCATION: (1230)..(1230)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1257)..(1257)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1265)..(1265)
OTHER INFORMATION: n equals a,t,g, or c
US-10-264-237-1192

Query Match 99.8%; Score 1078; DB 16; Length 1268;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1078; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATGCGCTGAGCTGAGCTGTGTCCCGGCGCTGGGTGGGGGCAACACCCGTCCTC 60
Db 121 ATGCGCTGAGCTGAGCTGTGTCCCGGCGCTGGGTGGGGGCAACACCCGTCCTC 180
Qy 61 ATCATTTGCCGAGATCGGCCAACAACAACGAGGAGCTGAGAGCTGACCAAGCTATGTC 120
Db 181 ATCATTTGCCGAGATCGGCCAACAACAACGAGGAGCTGAGAGCTGACCAAGCTATGTC 240
Qy 121 CGCATGCCCAAGAGTGTGGGGCTGATTTGGCAAGTTCAGAGAGTAGAGTGAATTC 180
Db 241 CGCATGCCCAAGAGTGTGGGGCTGATTTGGCAAGTTCAGAGAGTAGAGTGAATTC 300
Qy 181 AGCTTTATCGGAAAGCCTTGGAGAGGCAATACCTCGAAGCTTCTGGGGGAAGAG 240
Db 301 AGCTTTATCGGAAAGCCTTGGAGAGGCAATACCTCGAAGCTTCTGGGGGAAGAG 360
Qy 241 TAGGGGAGCAACAAGACATCTGAGCTCAGGCAAGCAAGACAGAGGAGTGCAGAG 300
Db 361 TAGGGGAGCAACAAGACATCTGAGCTCAGGCAAGCAAGACAGAGGAGTGCAGAG 420
Qy 301 TAGCGGAGAGGTGGGATCTTCTTCACTGCTCTGCAATGATGATGAGTGA 360
Db 421 TAGCGGAGAGGTGGGATCTTCTTCACTGCTCTGCAATGATGATGAGTGA 480
Qy 361 TTCTCGCATGAACTGAATGTTCCATTTTCAAGTTGATCTGGAGACCTAATATT 420
Db 481 TTCTCGCATGAACTGAATGTTCCATTTTCAAGTTGATCTGGAGACCTAATATT 540
Qy 481 TTCTCGCATGAACTGAATGTTCCATTTTCAAGTTGATCTGGAGACCTAATATT 540
Db 421 CCTTATCTGAAAAGACACCAAAAAGGTGCGCCCAATGCTGATCTCGATGAGATGCA 480
Qy 541 CCTTATCTGAAAAGACACCAAAAAGGTGCGCCCAATGCTGATCTCGATGAGATGCA 600
Db 481 TCAATGAGACCATGAGCAAGTTTATCAGATCTGTAGACCCCTTCAACCCCACTTCTGC 540
Qy 601 TCAATGAGACCATGAGCAAGTTTATCAGATCTGTAGACCCCTTCAACCCCACTTCTGC 660
Db 541 TTCTTCAGGTGACCAAGGCGATACCGCTCCAGCTGAGAGAGTCAACCTGCGGGTCATC 600
Qy 541 TTCTTCAGGTGACCAAGGCGATACCGCTCCAGCTGAGAGAGTCAACCTGCGGGTCATC 660
Db 661 TTCTTCAGGTGACCAAGGCGATACCGCTCCAGCTGAGAGAGTCAACCTGCGGGTCATC 720
Qy 601 TCGGAATATCAGAGCTCTTCTCTGACATTTCCATAGGGTATTCTGAGCAATGAAGGC 660
Db 721 TCGGAATATCAGAGCTCTTCTCTGACATTTCCATAGGGTATTCTGAGCAATGAAGGC 780
Qy 661 ATAGCGATATCTGTGCGCCGAGTGGCTGAGGGGCAAGGTGTGAGAGTCACTAACT 720
Db 781 ATAGCGATATCTGTGCGCCGAGTGGCTGAGGGGCAAGGTGTGAGAGTCACTAACT 840
Qy 721 TTGACACAGACCTGGAAGGGGAGTGAACCACTCGGCTCGCTGAGAGCTGGAAGTGGCC 780
Db 841 TTGACACAGACCTGGAAGGGGAGTGAACCACTCGGCTCGCTGAGAGCTGGAAGTGGCC 900
Qy 781 GAGCTGTGTGCGGTCAAGTGCCTTGTGTGAGCGTGCCTTGGGCTTCCCAACCAAGCACTG 840
Db 901 GAGCTGTGTGCGGTCAAGTGCCTTGTGTGAGCGTGCCTTGGGCTTCCCAACCAAGCACTG 960
Qy 841 CTGCTCTGTGAGATGCTGCTGATGAGAGCTGGGCAAGTCTGTGTGCGCAAGTGA 900

```
Db 961 CTGCGCTGTGAGTGGCTTGCATGAGAGCTGGGCAAGTCTGTGTGGCCAAAGTGA 1020
Qy 901 ATTCGGAAGGCAACATTTCTTAACATGACATGCTCAACCGTGAAGTGGTGAAGCCCAA 960
Db 1021 ATTCGGAAGGCAACATTTCTTAACATGACATGCTCAACCGTGAAGTGGTGAAGCCCAA 1080
Qy 961 GCGTATCTCTCGAAGACATCTTTATCTAGTGGGCAAGAGTCTGTGATCTTTGA 1020
Db 1081 GCGTATCTCTCGAAGACATCTTTATCTAGTGGGCAAGAGTCTGTGATCTTTGA 1140
Qy 1021 GAGGATGACACCATCATGAGAGAAATGTGATTAATCATGCAAAAAAATCAAGTCTTAA 1080
Db 1141 GAGGATGACACCATCATGAGAGAAATGTGATTAATCATGCAAAAAAATCAAGTCTTAA 1200

RESULT 4
US-10-037-270-1035
; Sequence 1035, Application US/10037270
; Publication No. US20030104529A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyun
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Weinman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yunging
; APPLICANT: Wang, Dantui
; APPLICANT: Wang, Zhiwei
; APPLICANT: Tillinghast, John
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. US20030104529A1el Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 784CIP2B
; CURRENT APPLICATION NUMBER: US/10/037,270
; CURRENT FILING DATE: 2002-01-04
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1104
; SOFTWARE: pt_PL_genes Version 1.0
; SEQ ID NO 1035
; LENGTH: 1230
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (119)..(1198)
US-10-037-270-1035

Query Match 99.7%; Score 1076.8; DB 15; Length 1230;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1078; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ATGCGCGTGTGAGTGGCTTGTGTCCCGGCGCTGGGTGGGCGGCAACCCGCTTC 60
Db 119 ATGCGCGTGTGAGTGGCTGTGTCCCGGCGCTGGGTGGGCGGCAACCCGCTTC 178
Qy 61 ATCATTTGCGAATCGGCGCAACCAACCGGCGAGCTGAGTCCCAAGCGCATGATC 120
Db 179 ATCATTTGCGAATCGGCGCAACCAACCGGCGAGCTGAGTCCCAAGCGCATGATC 238
Qy 121 CCGATGGCCAAAGAGTGTGGGCTGATTTGTGCCAAGTTCCAGAAAGTGTAGCATTC 180
Db 239 CCGATGGCCAAAGAGTGTGGGCTGATTTGTGCCAAGTTCCAGAAAGTGTAGCATTC 298
```

```
Qy 181 AAGTTATCGAAAGGCTTGGAGAGCCATACACCTCGAAGCATTCCTGTGGGGAAGAG 240
Db 299 AAGTTATCGAAAGGCTTGGAGAGCCATACACCTCGAAGCATTCCTGTGGGGAAGAG 358
Qy 241 TACCGGAGGCAAAACGACATCTGAGATTCAGCCATACCAAGTACAGGAGCTGCAGAG 300
Db 359 TACCGGAGGCAAAACGACATCTGAGATTCAGCCATACCAAGTACAGGAGCTGCAGAG 418
Qy 301 TACCGGAGGAGTGTGGATCTTCTTCACTGCTCTGCGCATAGATGAGATGAGCAAGTGA 360
Db 419 TACCGGAGGAGTGTGGATCTTCTTCACTGCTCTGCGCATAGATGAGATGAGCAAGTGA 478
Qy 361 TTCCGTGATGAACTGAATGTTTCAATTTTCAAAAGTGTGATGTGAGCACTAATATTT 420
Db 479 TTCCGTGATGAACTGAATGTTTCAATTTTCAAAAGTGTGATGTGAGCACTAATATTT 538
Qy 421 CCTATCTGAAAGGCAAGCCAAAGAGTGGCCCAATGGTATCTCCAGTGGATGAG 480
Db 539 CTTATCTGAAAGGCAAGCCAAAGAGTGGCCCAATGGTATCTCCAGTGGATGAG 598
Qy 481 TCAATGACACCATGAAAGAGTTTATCAGATGCTGAAGCCCTCAACCCCACTTCTGC 540
Db 599 TCAATGACACCATGAAAGAGTTTATCAGATGCTGAAGCCCTCAACCCCACTTCTGC 658
Qy 541 TTCTTGCAAGTGTACCAAGGCAATCCGCTCAAGCTGAGAGCTGAGAGCTGAGTATC 600
Db 659 TTCTTGCAAGTGTACCAAGGCAATCCGCTCAAGCTGAGAGCTGAGAGCTGAGTATC 718
Qy 601 TCGGAATATCGAAGCTTTTCTTCTGACATTCCTCAATGAGGTATTTCTGGGCAAGAAAG 660
Db 719 TCGGAATATCGAAGCTTTTCTTCTGACATTCCTCAATGAGGTATTTCTGGGCAAGAAAG 778
Qy 661 ATAGCATATCTGTGGCCGCAAGTGTCTGTGGGCAAGGTGTGGAAGCTGCATTAAT 720
Db 779 ATAGCATATCTGTGGCCGCAAGTGTCTGTGGGCAAGGTGTGGAAGCTGCATTAAT 838
Qy 721 TTGGAAGAAGCTTGAAGGAGGAGTGAACCACTGGGCTGTGAGGCTGAGAACTGGCC 780
Db 839 TTGGAAGAAGCTTGAAGGAGGAGTGAACCACTGGGCTGTGAGGCTGAGAACTGGCC 898
Qy 781 GAGTGTGTGGTGTAGTGGCTTGTGAGAGCTGCTGTGGGCTGTGAGGCTGTGAGGCTG 840
Db 899 GAGTGTGTGGTGTAGTGGCTTGTGAGAGCTGCTGTGGGCTGTGAGGCTGTGAGGCTG 958
Qy 841 CTGCGCTGTGAGTGGCTGTGAGAGCTGAGAGCTGTGAGTGTGAGTGTGAGTGTGAG 900
Db 959 CTGCGCTGTGAGTGGCTGTGAGAGCTGAGAGCTGTGAGTGTGAGTGTGAGTGTGAG 1018
Qy 901 ATTCGGAAGGCAACATTTCTTAACATGACATGCTCAACCGTGAAGTGGTGAAGCCCAA 960
Db 1019 ATTCGGAAGGCAACATTTCTTAACATGACATGCTCAACCGTGAAGTGGTGAAGCCCAA 1078
Qy 961 GCGTATCTCTCGAAGACATCTTTATCTAGTGGGCAAGAGTCTGTGATCTTTGA 1020
Db 1079 GCGTATCTCTCGAAGACATCTTTATCTAGTGGGCAAGAGTCTGTGATCTTTGA 1138
Qy 1021 GAGGATGACACCATCATGAGAGAAATGTGATTAATCATGCAAAAAAATCAAGTCTTAA 1080
Db 1139 GAGGATGACACCATCATGAGAGAAATGTGATTAATCATGCAAAAAAATCAAGTCTTAA 1198

RESULT 5
US-10-117-722-1035
; Sequence 1035, Application US/10117722
; Publication No. US2003021974A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. US2003021974A1el Nucleic Acids and
```

TITLE OF INVENTION: Polypeptides
 FILE REFERENCE: 784CIP2BCIP
 CURRENT APPLICATION NUMBER: US/10/117,722
 CURRENT FILING DATE: 2002-04-04
 PRIOR APPLICATION NUMBER: 09/620,312
 PRIOR FILING DATE: 2000-07-19
 PRIOR APPLICATION NUMBER: 09/552,317
 PRIOR FILING DATE: 2000-04-25
 PRIOR APPLICATION NUMBER: 09/498,725
 PRIOR FILING DATE: 2000-01-21
 NUMBER OF SEQ ID NOS: 1104
 SOFTWARE: Pf. PL_genes Version 1.0
 SEQ ID NO 1035
 LENGTH: 1230
 TYPE: DNA
 ORGANISM: Homo sapiens
 FEATURE:
 NAME/KEY: CDS
 LOCATION: (119)..(1198)
 US-10-117-722-1035

Query Match 99.7%; Score 1076.8; DB 16; Length 1230;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 1076; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGCCCCGTGAGCTGAGCTGTCTCCCGGGCGCTGGGTGGCGGGGCAACACCCGCTCTC 60
 DB 119 ATGCCCCGTGAGCTGAGCTGTCTCCCGGGCGCTGGGTGGCGGGGCAACACCCGCTCTC 178
 QY 61 ATCATTTGCCAGATCGCGCGAGAACCAACAGAGGCGACCTGAGACCTAGCCAGCCATGATC 120
 DB 179 ATCATTTGCCAGATCGCGCGAGAACCAACAGAGGCGACCTGAGACCTAGCCAGCCATGATC 238
 QY 121 CGCATGGCCAGAGAGTGTGGGGCTGATTTGGCCAACTTCCAGAGAGTGAAGCTGAATTC 180
 DB 239 CGCATGGCCAGAGAGTGTGGGGCTGATTTGGCCAACTTCCAGAGAGTGAAGCTGAATTC 298
 QY 181 AAGTTTAAATCGGAAAGCTTGGAGAGGCGCATACACCTCGAAGATTCCTGGGGGAGAGAG 240
 DB 299 AAGTTTAAATCGGAAAGCTTGGAGAGGCGCATACACCTCGAAGATTCCTGGGGGAGAGAG 358
 QY 241 TACGGGAGCAAAACGACATCTGGAGTTGAGCCATGACAGCAATGACAGGAGCTGACAGAG 300
 DB 359 TACGGGAGCAAAACGACATCTGGAGTTGAGCCATGACAGCAATGACAGGAGCTGACAGAG 418
 QY 301 TACGGGAGAGAGTGGAGATCTTCACTGCGCTGCGCATGAGATGAGATGAGAGTTGAA 360
 DB 419 TACGGGAGAGAGTGGAGATCTTCACTGCGCTGCGCATGAGATGAGATGAGAGTTGAA 478
 QY 361 TTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
 DB 479 TTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 538
 QY 421 CTTTATTTGAAAGAGACCCAAAGAGTCCCGCCCAATGATGATGATGATGATGATGATGATG 480
 DB 539 CTTTATTTGAAAGAGACCCAAAGAGTCCCGCCCAATGATGATGATGATGATGATGATGATG 598
 QY 481 TCAATGAGCAACATGAGCAAGTTTATCAGATGATGAGAGCCCTCAACCCCACTTCTGC 540
 DB 599 TCAATGAGCAACATGAGCAAGTTTATCAGATGATGAGAGCCCTCAACCCCACTTCTGC 658
 QY 541 TTTCTGAGTGTACACCGCATACCCGCTTCAAGCTGAGAGCTCAAGCTGAGAGCTGAGT 600
 DB 659 TTTCTGAGTGTACACCGCATACCCGCTTCAAGCTGAGAGCTCAAGCTGAGAGCTGAGT 718
 QY 601 TCGGAATATCAGAGCTCTTCTGACATTCATAGGATGATGATGATGATGATGATGATGATG 660
 DB 719 TCGGAATATCAGAGCTCTTCTGACATTCATAGGATGATGATGATGATGATGATGATGATG 778
 QY 661 ATAGCGATATCTGTGGCGGAGTGTCTTGGGGGCGCAAGGTGTTGAGAGCTCAATTAAT 720
 DB 779 ATAGCGATATCTGTGGCGGAGTGTCTTGGGGGCGCAAGGTGTTGAGAGCTCAATTAAT 838

QY 721 TTGACACAGACTTGAGAGGGAGGTGACCACTTGCGCTGCTTGGAGCTTGAGAACTGGCC 780
 DB 839 TTGACACAGACTTGAGAGGGAGGTGACCACTTGCGCTGCTTGGAGCTTGAGAACTGGCC 898
 QY 781 GAGCTGTGCGGCTGAGTGTGTGTGAGAGCTGCGCTTGGGCTCTCCCAACCAAGCAGCTG 840
 DB 899 GAGCTGTGCGGCTGAGTGTGTGTGAGAGCTGCGCTTGGGCTCTCCCAACCAAGCAGCTG 958
 QY 841 CTGCGCTGTGAGATGAGCTTGTGATGAGAACTGGGCAAGTCTGTGTGCGCAAGTGAAA 900
 DB 959 CTGCGCTGTGAGATGAGCTTGTGATGAGAACTGGGCAAGTCTGTGTGCGCAAGTGAAA 1018
 QY 901 ATTCCGGAAGGACCAATCTTACATGAGACAGCTCAACCGTGAAGTGGGTGAGCCCAA 960
 DB 1019 ATTCCGGAAGGACCAATCTTACATGAGACAGCTCAACCGTGAAGTGGGTGAGCCCAA 1078
 QY 961 GCGTATCTCTCTGAGAGACATCTTTATATCTAGTGGGCAAGAGTCTGTGCACTGTGAA 1020
 DB 1079 GCGTATCTCTCTGAGAGACATCTTTATATCTAGTGGGCAAGAGTCTGTGCACTGTGAA 1138
 QY 1021 GAGGATGACCATATGAGAAATTTGTAGTATGATGAGCAAAATTAAGTCTTAA 1080
 DB 1139 GAGGATGACCATATGAGAAATTTGTAGTATGATGAGCAAAATTAAGTCTTAA 1198

RESULT 6
 US-09-880-107-1514/C
 Sequence 1514; Application US/09880107
 Patent No. US20020142981A1
 GENERAL INFORMATION:
 APPLICANT: Horne, Darci T.
 APPLICANT: Vockley, Joseph G.
 APPLICANT: Scherf, Iwe
 APPLICANT: Gene Logic, Inc.
 TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
 FILE REFERENCE: 44921-5028-WO
 CURRENT APPLICATION NUMBER: US/09/880,107
 CURRENT FILING DATE: 2001-06-14
 PRIOR APPLICATION NUMBER: US 60/211,379
 PRIOR FILING DATE: 2000-06-14
 PRIOR APPLICATION NUMBER: US 60/237,054
 PRIOR FILING DATE: 2000-10-02
 NUMBER OF SEQ ID NOS: 3950
 SOFTWARE: Patent In Ver. 2.1
 SEQ ID NO 1514
 LENGTH: 464
 TYPE: DNA
 ORGANISM: Homo sapiens
 OTHER INFORMATION: Genbank Accession No. US20020142981A1 AA621146
 US-09-880-107-1514

Query Match 38.7%; Score 418; DB 9; Length 464;
 Best Local Similarity 98.6%; Pred No. 4; 4e-124;
 Matches 432; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

QY 643 TCTGGCATGAAACAGGATAGCGATATCTGTGCGCGAGTGGCTTGGGGGCGCAAGTGT 702
 DB 464 TCTGGCATGAAACAGGATAGCGATATCTGTGCGCGAGTGGCTTGGGGGCGCAAGTGT 405
 QY 703 TTGGAACGTCATTAATCTTGGACAAAGCTGGAAGGGAGTGAACAAGTGGCTGCTG 762
 DB 404 TTGGAACGTCATTAATCTTGGACAAAGCTGGAAGGGAGTGAACAAGTGGCTGCTG 345
 QY 763 GAGCTTGAGAACTGGCGGAGCTGTGTGAGTGTCTTGTGAGAGCTGCTGCTGCTG 822
 DB 344 GAGCTTGAGAACTGGCGGAGCTGTGTGAGTGTCTTGTGAGAGCTGCTGCTGCTG 285
 QY 822 TCCCAACCAAGCAGCTGTGCGCTGTGAGATGAGCTGCAATGAGAAAGTGGCAAGTCT 882
 DB 284 TCCCAACCAAGCAGCTGTGCGCTGTGAGATGAGCTGCAATGAGAAAGTGGCAAGTCT 225
 QY 883 GTGTGGCCCAAGTGAATAATTCGGAAGGCAACATTTTAAACATGAGCAATGCTCAACCGTG 942

```
Db 224 GTGGTGGCCAAAGTGAATATCCGGAAGCAACATTTACATGACATGCTCAACGCTG 165
Qy 943 AAGGTGGGTGAGCCCAAGCCATCTCTCTGTAAGCATCTTTAATCTGTGGCAAGAG 1002
Db 164 AAGGTGGGTGA-CCCAAGGCTATCTCTCTGTAAGCATCTTTAATCTGTGGCAAGAG 106
Qy 1003 GTCCGTGCTCACTGTGTAAGAGATGACACCATCATGTAAGAATTGGTATATCATGGC 1062
Db 105 GTCCGTGCTCACTGTGTAAGAGATGACACCATCATGTAAGAATTGGTATATCATGGC 46
Qy 1063 AAAAAATCAAGTCTTAA 1080
Db 45 AAAAAATCAAGTCTTAA 28

RESULT 7
US-09-918-995-5220
; Sequence 5220, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5220
; LENGTH: 486
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(486)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-5220

Query Match 34.4%; Score 371.8; DB 10; Length 486;
Best Local Similarity 99.5%; Pred. No. 3.9e-109;
Matches 373; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AATCCGCTGAGCTGAGAGCTGTGTCCCGGCGCTGGTGGGCGGCAACCCGCTTC 60
Db 111 ATGCGCTGAGCTGAGCTGTGTCCCGGCGCTGGTGGGCGGCAACCCGCTTC 170
Qy 61 ATCATTTGCCGAATCGGCGCAACCAACAGGCGACCTGAGCGTACGCAAGCGCATATC 120
Db 171 ATCATTTGCCGAATCGGCGCAACCAACAGGCGACCTGAGCGTACGCAAGCGCATATC 230
Qy 121 CGCATGGCCAAAGAGTGTGGGGCTGATTTGTCCCAAGTTCAGAAAGTGAAGTAAATTC 180
Db 231 CGCATGGCCAAAGAGTGTGGGGCTGATTTGTCCCAAGTTCAGAAAGTGAAGTAAATTC 290
Qy 181 AAGTTAATCGAAAGACCTTGGAGAGGCGATACACCTCGAAGCATTTCTGGGGCAAGAG 240
Db 291 AAGTTAATCGAAAGACCTTGGAGAGGCGATACACCTCGAAGCATTTCTGGGGCAAGAG 350
Qy 241 TACGGGAGACCAACGACATCTGAGTTCAAGCCATGACCATGACGAGAGCTGCAGAG 300
Db 351 TACGGGAGACCAACGACATCTGAGTTCAAGCCATGACCATGACGAGAGCTGCAGAG 410
Qy 301 TACGGGAGAGAGTGGAGTCTTCTCACTGCTCTGGCATGAGATGAGAGTGA 360
Db 411 TACGGGAGAGAGTGGAGTCTTCTCACTGCTCTGGCATGAGATGAGAGTGA 470
Qy 361 TTCTCTCATGACTG 375
Db 471 TTCTCTCATGACTG 485
```

```
RESULT 8
US-09-918-995-28620
; Sequence 28620, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 28620
; LENGTH: 442
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(442)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-28620

Query Match 31.2%; Score 336.8; DB 10; Length 442;
Best Local Similarity 96.0%; Pred. No. 7.6e-98;
Matches 341; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 2 TGCCGTGAGCTGAGAGCTGTGTCCCGGCGCTGGTGGGCGGCAACACCGTCTTCA 61
Db 90 TGCCGTGAGCTGAGAGCTGTGTCCCGGCGCTGGTGGGCGGCAACACCGTCTTCA 149
Qy 62 TCATTGCCGAGATGGCCCAAGCCACGACGAGGCGACCTGAGCGTACGAGCGATATCC 121
Db 150 TCATTGCCGAGATGGCCCAAGCCACGACGAGGCGACCTGAGCGTACGAGCGATATCC 209
Qy 122 GCATGGCCAAAGAGTGTGGGGCTGATTTGTCCCAAGTTCAGAAAGTGAAGTAAATTC 181
Db 210 GCATGGCCAAAGAGTGTGGGGCTGATTTGTCCCAAGTTCAGAAAGTGAAGTAAATTC 269
Qy 182 AGTTAATCGAAAGACCTTGGAGAGGCGATACACCTCGAAGCATTTCTGGGGCAAGAG 241
Db 270 AGTTAATCGAAAGACCTTGGAGAGGCGATACACCTCGAAGCATTTCTGGGGCAAGAG 329
Qy 242 ACGGGAGACCAACGACATCTGAGTTCAAGCCATGACCATGACGAGAGCTGCAGAG 301
Db 330 ACGGGAGACCAACGACATCTGAGTTCAAGCCATGACCATGACGAGAGCTGCAGAG 389
Qy 302 ACGGGAGAGAGTGGAGTCTTCTCACTGCTCTGGCATGAGATGAGAGTGA 349
Db 390 ACGGGAGAGAGTGGAGTCTTCTCACTGCTCTGGCATGAGATGAGAGTGA 437
```

```
RESULT 9
US-09-918-995-25343
; Sequence 25343, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 25343
; LENGTH: 417
; TYPE: DNA
```



```

APPLICANT: Ruben et al.
TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
FILE REFERENCE: PA005PI
CURRENT APPLICATION NUMBER: US/10/106,698
CURRENT FILING DATE: 2002-03-27
PRIOR APPLICATION NUMBER: PCT/US00/26524
PRIOR FILING DATE: 2000-09-28
PRIOR APPLICATION NUMBER: US 60/157,137
PRIOR FILING DATE: 1999-09-29
PRIOR APPLICATION NUMBER: US 60/163,280
PRIOR FILING DATE: 1999-11-03
NUMBER OF SEQ ID NOS: 8564
SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 2649
LENGTH: 410
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (287)..(287)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (300)..(300)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (370)..(370)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (374)..(374)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (402)..(402)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (405)..(405)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (408)..(408)
OTHER INFORMATION: n equals a,t,g, or c
US-10-106-698-2649

```

```

Query Match      23.7% Score 256.2; DB 15; Length 410;
Best Local Similarity 97.2%; Pred. No. 8.2e-72;
Matches 280; Conservative 0; Mismatches 6; Indels 2; Gaps 2;

```

```

QY 548 AGGTACCAAGGAGATCCGCTCCAGCTGAGAGAGCAACCTGCGGTCATCTCGAAT 607
DB 110 AGGTACCAAGGAGATCCGCTCCAGCTGAGAGAGCAACCTGCGGTCATCTCGAAT 169
QY 608 ATCAGAGCTCTTCTGACATTCATAGGATTTCTGGGCAATGAAACAGCATAGCA 657
DB 170 ATCAGAGCTCTTCTGACATTCATAGGATTTCTGGGCAATGAAACAGCATAGCA 229
QY 668 TATCTGTGGCGAGAGTGGCTCTGGGAGGCAAGGTGTAAGCACTTAACCTTGACA 727
DB 230 TATCTGTGGCGAGAGTGGCTCTGGGAGGCAAGGTGTAAGCACTTAACCTTGACA 289
QY 728 AGACTGGAAGGGAGAGTGAACACTCGGCTCGCTGAGAGCTGAGAACTGGCCG-AGCTG 786
DB 290 AGACTGGAAGGGAGAGTGAACACTCGGCTCGCTGAGAGCTGAGAACTGGCCGAGCTG 349
QY 787 GTGCGGTCAATGTGTCTGTGT-AGAGAGTGGCTGGGCTCCCAAGCAA 833
DB 350 GTGCGGTCAATGTGTCTGTGTGTGAGAGCTGCGCTGGGCTCCCAAGCAA 397

```

```

RESULT 13
US-10-085-783A-31232
Sequence 31232, Application US/10085783A
Publication No. US20040037841A1
GENERAL INFORMATION:
APPLICANT: ChondroGene Inc.
APPLICANT: Liew, C.C.

```

```

TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
FILE REFERENCE: 4231/2002
CURRENT APPLICATION NUMBER: US/10/085,783A
CURRENT FILING DATE: 2002-02-28
PRIOR APPLICATION NUMBER: US 60/305,340
PRIOR FILING DATE: 2001-07-13
PRIOR APPLICATION NUMBER: US 60/275,017
PRIOR FILING DATE: 2001-03-12
PRIOR APPLICATION NUMBER: US 60/271,955
PRIOR FILING DATE: 2001-02-28
NUMBER OF SEQ ID NOS: 58994
SOFTWARE: PatentIn version 3.2
SEQ ID NO 31232
LENGTH: 203
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc feature
LOCATION: (6)..(6)
OTHER INFORMATION: n is a, c, g, or t
NAME/KEY: misc feature
LOCATION: (31)..(31)
OTHER INFORMATION: n is a, c, g, or t
US-10-085-783A-31232

```

```

Query Match      16.7% Score 180.4; DB 13; Length 203;
Best Local Similarity 98.4%; Pred. No. 1.7e-47;
Matches 181; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 897 GAAATTCGGAAGGACCATTTCTACATGACATGCTACCGTGAAAGTGGGAGACC 956
DB 1 GAAATTCGGAAGGACCATTTCTACATGACATGCTACCGTGAAAGTGGGAGACC 60
QY 957 CAAGCCTATCCTCTGGAAGCATCTTATCTAGTGGCAAGAGTCTGTCACTGT 1016
DB 61 CAAGCCTATCCTCTGGAAGCATCTTATCTAGTGGCAAGAGTCTGTCACTGT 120
QY 1017 TGAAGAGATGACACCATCATGAGAAATTTGATATCATGCAAAAATTAATC 1076
DB 121 TGAAGAGATGACACCATCATGAGAAATTTGATATCATGCAAAAATTAATC 180
QY 1077 TTTA 1080
DB 181 TTTA 184

```

```

RESULT 14
US-10-242-535A-31232
Sequence 31232, Application US/10242535A
Publication No. US2004001363A1
GENERAL INFORMATION:
APPLICANT: ChondroGene Inc.
APPLICANT: Liew, C.C.
TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
FILE REFERENCE: 4231/2005
CURRENT APPLICATION NUMBER: US/10/242,535A
CURRENT FILING DATE: 2002-09-12
PRIOR APPLICATION NUMBER: US 10/085,783
PRIOR FILING DATE: 2002-02-28
PRIOR APPLICATION NUMBER: US 60/305,340
PRIOR FILING DATE: 2001-07-13
PRIOR APPLICATION NUMBER: US 60/275,017
PRIOR FILING DATE: 2001-03-12
PRIOR APPLICATION NUMBER: US 60/271,955
PRIOR FILING DATE: 2001-02-28
NUMBER OF SEQ ID NOS: 58994
SOFTWARE: PatentIn version 3.2
SEQ ID NO 31232
LENGTH: 203
TYPE: DNA
ORGANISM: Human
FEATURE:

```

Thu May 13 11:53:07 2004

US-09-930-440b-5.mprb

Page 9

NAME/KEY: misc_feature
LOCATION: (1)..(6)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (31)..(21)
OTHER INFORMATION: n is a, c, g, or t
US-10-242-555A-31222

Query Match	16.7%;	Score 180.4;	DB 16;	Length 203;
Best Local Similarity	98.4%;	Pred. No. 1.7e-47;		
Matches 181;	Conservative 0;	Mismatches 3;	Indels 0;	Gaps 0

QY	897	GAAATTCGGGAAGGCACCATTTCAACATGGACATGCTCACCGTGAAGAGTGGGTGACC	956
Db	1	GAAANTCGGGAAGGCACCATTTCTAACAAATNACATGCTCACCGTGAAGAGTGGGTGACC	60
QY	957	CAAAECTATCCTCCTGAAAGACATTTTAAATTAAGTGGGCAAGAAAGTCTGCTACGT	1018
Db	61	CAAAAGCTATCCTCCTGAAAGACATTTTAAATTAAGTGGGCAAGAAAGTCTGCTACGT	120
QY	1017	TGAAGAGATGACACCATTCATCGAAGAAATGGTAAATATATCATGGCAAAAAAATCAATTC	1076
Db	121	TGAAGAGATGACACCATTCATCGAAGAAATGGTAAATATATCATGGCAAAAAAATCAAGTTC	180
QY	1077	TTAA 1080	
Db	181	TTAA 184	

RESULT 15
US-10-156-761-3360
; Sequence 3360, Application US/10156761
; Publication No. US20030119018A1

```

1  APPLICANT:  OMURA, SATOSHI
2  APPLICANT:  IKEYA, HARUO
3  APPLICANT:  ISHIZAWA, JUN
4  APPLICANT:  HORIKAWA, HIROSHI
5  APPLICANT:  SHIBA, TADAYOSHI
6  APPLICANT:  SAKAKI, YOSHIYUKI
7  APPLICANT:  HATTORI, MASAHIRA
8  TITLE OF INVENTION:  NOVEL POLYNUCLEOTIDES
9  FILE REFERENCE:  249-262
10 CURRENT APPLICATION NUMBER:  US/10/156, 761
11 CURRENT FILING DATE:  2002-05-29
12 PRIOR APPLICATION NUMBER:  JP 2001-204089
13 PRIOR FILING DATE:  2001-05-30
14 PRIOR APPLICATION NUMBER:  JP 2001-272697
15 PRIOR FILING DATE:  2001-06-02
16 NUMBER OF SEQ ID NOS:  15109
17 SEQ ID NO 3360
18 LENGTH:  933
19 TYPE:  DNA
20 ORGANISM:  Streptomyces avermitilis
21 FEATURE:
22 NAME/KEY:  CDS
23 LOCATION:  (1)..(933)
24
25 US-10-156-761-3360

```

Query Match	12.5%;	Score 135;	DB 15;	Length 933;
Best Local Similarity	49.3%;	Pred. NO. 2e-32;		
Matches 413; Conservative	0;	Mismatches 415;	Indels 9;	Gaps 2

QY 40 GGCGGGCAACCCGCTCTTCATCATTTGGCCGAGATGCGCAGAACCAACGCGGACCTG 99
Db 43 GGCCCGGGCCACCCCGCTACGTCGTCGGGAGATGGATCAACCAACGCGAGCTC 1028
QY 100 GACCTACCAAGCCCAATGATCCGCAATGGCCAGAGATGTGGGCTGATTTGGCCAAATTC 159
Db 103 GAGAAACGCTTCAAGCTATATGACGCCCGCCCGCAGACCCGCTGCGAGCGCCCTCAATTC 1628
QY 160 CAGAAAGTGAAGCTAGATTCAGATTAAATCGGAAAGCCTTGGAGAGCCATACACTTCG 218

Db	163	CAGAAACGCAACCCCGGAGATCTGCACCCCGGCGACCAAGTGGAGCATTCAGCGCGACACC	222
Qy	220	AAGCATTTCTGGGGGGAAGACGTACCGGGGAGAGCAAAACGACATCTGGAGTTCAAGCATAGAC	275
Db	223	---CCCTGGGGGCGGATATCACTACATGCATACCGGCACCGCGTGGAGTTGGTGGAGAAC	279
Qy	280	CAGTACAGGAGACCTGCAGAGGTACGCGCCGAGAGAGGTGGGAATCTTTCTTCACTGCCTTGGC	335
Db	280	GAGTACCGGCGCAGATGCAGAGTACGCCCAAGAGCAAAACATTCGACTGGTTGGCTCTCCCG	335
Qy	340	ATGGAGATGAATGGCACTTGAATTTCTGTGCATGACAGTATTCATTTTCAAGTTTGA	399
Db	340	TGGAGACCGAGGCGGTGCGCTTCTCTGGAGAGTTTGCACATCCCGGCCCAAGAGTGGCC	399
Qy	400	TCTGGAGACATATATATTTTCTTATCTGMAAAAGACAGCGCAAAAAGTGGCCCAATG	455
Db	400	TCCGGGTCTCTGACCGAGCAGACGAGGTGCTCGCGGCTCTGCGCGGTACCGGCGCAGCGTTC	455
Qy	460	GTGATTTCCAGTGGAGTGCAGTGCATATGACACACATGAAGCAAGTTTATCAATGTGTAG	515
Db	460	ATCTCTTCCAGGGGCAATGTGCAC-----CCGAAAGCAGATCCGTCAAGCGGTGAGGTC	515
Qy	520	CCCCTCAACCCCAACTTCTGCTTCTTTCGACGTATCACAGCGATACCGGCTCAGACTTGAG	575
Db	514	CTCGGGCTCGGACAAACATCTCGCTCTGACACGCGACCTGAGTATCCGCGCAGACGCGGAG	575
Qy	580	GAGCTAACCTGGGGTCAATCTCGGATATATCGAAGAGCTCTTTCGACATTTCCATGTAGG	635
Db	574	GACCTCAACTCTCGGCTCATCAACACCTCCAGCGCGAGTACCAGACGTCCCATGAGC	635
Qy	640	TATTTCTGGGATGAAACAGGCAATAGCGATATCTGTGGCGCGAGTGGCTTGGGGGCGAAG	699
Db	634	TATCTCCGGCCACGAGCGGGCTGTGCAGACCAAGCGCTGCGGGGTGCGCTCTCGGCGCAG	699
Qy	700	GTTGTGGAGAGTACATTAATCTTTGGACAAGACCTGGAGAGGGGAGTGAACACTCGGCGCTGG	755
Db	694	TTTGTCGAGGCGACATCAACCTTCGACGCGCGCAATGTGGGCTCTGACCAAGCCGCTTCC	755
Qy	760	CTGAGACTTGGAGAACTGGCGCAGAGCTGTGCGGTCAATGTCGTTTGTGAAGCTGTGCTTG	815
Db	754	GTCGAGCGGACGGGCTTCAACGCGCTGTGCTGTGATCATTCGACACCATGAGAGCGTCTCTC	815
Qy	820	GGCTCCCAACCAAGCAGCTGCGCCCTGTGAGATGAGGCTGCAATGAGACACTGGGC	875
Db	814	GCGAGCGGCTTCAAGAGGTTATCAAGTTCGAGCTCGGCGCCCATGAAAGAACTGGC	870

Search completed: May 12, 2004, 12:37:27
Job time : 508.439 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 08:55:49 ; Search time 50.614 Seconds
(without alignments)
2004.184 Million cell updates/sec

Title: US-09-930-440b-6
Perfect score: 1887

Sequence: 1 MPLELELCPRKRWGQGHPCF.....EEDDTIMEELVDNHKKIKS 359

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04.*

1: geneseqp1980s.*
2: geneseqp1990s.*
3: geneseqp2000s.*
4: geneseqp2001s.*
5: geneseqp2002s.*
6: geneseqp2003as.*
7: geneseqp2003bs.*
8: geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1887	100.0	359	3 AAY96101	AAY96101 Human sla
2	1887	100.0	359	3 AAY90352	AAY90352 Human gly
3	1887	100.0	359	4 AAB84683	AAB84683 Amino aci
4	1887	100.0	359	6 AAO26546	AAO26546 Human SA
5	1883	99.8	359	5 ABB90221	ABB90221 Human pol
6	1880	99.6	359	4 AAM39986	AAM39986 Human pro
7	1876	99.4	359	4 AAB93183	AAB93183 Human pro
8	1789	94.8	400	4 AAM41772	AAM41772 Human pol
9	783	41.5	338	4 ABB71364	ABB71364 Drosophil
10	544.5	28.9	338	3 AAY68965	AAY68965 Cp82P pro
11	532.5	28.2	341	5 AHP26810	AHP26810 Streptococ
12	510.5	27.1	346	3 AAY96102	AAY96102 Escherich
13	510.5	27.1	346	6 AAB84684	AAB84684 Amino aci
14	510.5	27.1	346	6 AAO26547	AAO26547 Bacterial
15	419.5	22.2	125	4 ABB17251	ABB17251 Novel hum
16	395	20.9	153	4 ABB27650	ABB27650 Novel hum
17	382.5	20.3	349	5 AAM49716	AAM49716 N. mening
18	381	20.2	123	3 AAG75152	AAG75152 Human col
19	336.5	17.8	346	3 AAY97207	AAY97207 Campyloba
20	336.5	17.8	346	6 AAB18497	AAB18497 Campyloba
21	320.5	17.0	348	6 ABB78189	ABB78189 Amino aci
22	303	16.1	340	2 AAW98812	AAW98812 H. pylori
23	266.5	14.1	136	4 ABB27649	ABB27649 Novel hum
24	159.5	8.5	201	5 ABB51188	ABB51188 Helicobac
25	149.5	7.9	182	5 ABB50662	ABB50662 Helicobac

26	113.5	6.0	333	6 ABB29572	ABB29572 Protein e
27	113.5	6.0	343	7 ADC97084	ADC97084 B. faeciu
28	110	5.8	265	4 AAB96139	AAB96139 Putative
29	108.5	5.7	633	6 ABB35582	ABB35582 Protein e
30	107.5	5.7	337	6 ABB23592	ABB23592 Protein e
31	107	5.7	310	6 ABB39673	ABB39673 Protein e
32	104.5	5.5	769	6 ABB70636	ABB70636 Phototrab
33	102.5	5.4	610	6 ABB06924	ABB06924 Altolococ
34	102.5	5.4	629	6 ABB06926	ABB06926 Altolococ
35	102.5	5.4	630	6 ABB06928	ABB06928 Altolococ
36	101.5	5.4	337	6 ABB25380	ABB25380 Protein e
37	101.5	5.4	341	6 ABB29304	ABB29304 Protein e
38	101	5.4	66	7 ADC79285	ADC79285 Z. elonga
39	101	5.4	350	7 ADC59279	ADC59279 Comamonas
40	101	5.4	350	7 ADC52503	ADC52503 Testoster
41	100	5.3	65	7 ADC79293	ADC79293 Z. elonga
42	100	5.3	547	6 ADA12178	ADA12178 Actinobac
43	100	5.3	547	7 ADC61133	ADC61133 Baeyer-Vi
44	100	5.3	547	7 ABB39407	ABB39407 Actinobac
45	100	5.3	1109	6 ABB71017	ABB71017 Human adi

ALIGNMENTS

RESULT 1
ID AAY96101 standard; protein, 359 AA.
XX AAY96101;
AC AAY96101;
DT 19-DEC-2000 (first entry)
XX
DE Human sialic acid synthetase.
XX
FM Sialic acid synthetase; human; sas gene; sialylation; glycoprotein;
KM plasmidogen; transferin; thyrotropin; Nat,K-ATPase.
XX
OS Homo sapiens.
XX
PN WC0200052135-A2.
XX
PD 08-SEP-2000.
XX
PF 01-MAR-2000; 2000MO-US005313.
XX
PR 02-MAR-1999; 99US-0122582P.
PR 08-DEC-1999; 99US-0169624P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (UYJO) UNIV JOHN HOPKINS.
PA (UYWY-) UNIV WYOMING.
XX
PI Betenbaugh MJ, Lawrence S, Lee YC, Jarvis D, Coleman TR,
XX MPI, 2000-572178/53.
DR N-PSDB; AAA50569.
XX
PT Recombinant production of sialylated glycoproteins using cells in which
PT the expression of enzymes, e.g. sialic acid synthetase, involved in the
PT sialylation reaction has been altered.
XX
PS Claim 16; Page 105-106; 144pp; English.
XX
CC The present sequence is that of human sialic acid synthetase (SAS), an
CC enzyme that condenses Mann6-P or Man-6-P with GDP to form Neu5Ac and
CC KMN phosphates, respectively. The sequence was deduced from SAS cDNA (see
CC AA050569). Northern blots indicated ubiquitous transcription of the SAS
CC gene in a selection of tissues. The invention provides methods and
CC recombinantly engineered cells for producing glycoproteins having
CC sialylated oligosaccharides. The methods involve altering the expression
CC of enzymes involved in carbohydrate processing. A claimed cell producing
CC sialylated glycoprotein at above endogenous levels expresses at least 1

RESULT 3
 AAB84683
 ID AAB84683 standard; protein; 359 AA.
 XX
 AC AAB84683;
 XX
 DT 17-SEP-2001 (first entry)
 XX
 DE Amino acid sequence of a human sialic acid synthetase.
 XX
 KM Carbohydrate processing; sialylation; sialylated glycoprotein; CMP-SA;
 KM Cytidine monophosphate-sialic acid; N-acetylglucosamine-2 epimerase;
 KM Sialic acid synthetase; aldolase; CMP-SA synthetase; CMP-SA transporter;
 KM vaccine.
 XX
 OS Homo sapiens.
 XX
 PN WO200142492-A1.
 PD 14-JUN-2001.
 XX
 PF 07-DEC-2000; 2000MO-US033136.
 XX
 PR 09-DEC-1999; 99US-0169839P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 PA (UYJO) UNIV JOHNS HOPKINS.
 PA (UTEM) UNIV TEMPLE.
 PA (UYWY-) UNIV WYOMING.
 XX
 PI Betenbaugh MJ, Lawrence S, Lee YC, Coleman TA, Palter K;
 PI Jarvis D;
 PI WPI; 2001-441575/47.
 DR N-PSD3; AAB28458.
 XX
 PT Cells producing cytidine monophosphate-sialic acid and sialylated
 PT glycoprotein above endogenous levels for production of vaccines and
 PT therapeutics.
 PS
 PS Claim 16; Fig 32; 182pp; English.
 XX
 CC The specification describes a method for manipulating carbohydrate
 CC processing pathways in cells of interest. The methods are used to
 CC manipulate multiple pathways involved with the sialylation reaction by
 CC using recombinant DNA technology and substrate feeding approaches to
 CC enable the production of sialylated glycoproteins in the cells. The
 CC sialylation process involves the post-translational addition of the donor
 CC substrate cytidine monophosphate-sialic acid (CMP-SA) onto a specific
 CC acceptor carbohydrate. The cells express at least one enzyme, selected
 CC from N-acetylglucosamine-2 epimerase, sialic acid synthetase, aldolase,
 CC CMP-SA synthetase and CMP-SA transporter, above endogenous levels. The
 CC cells are useful for producing complex sialylated glycoproteins in cells
 CC of interest, especially insect cells. Glycoproteins containing sialylated
 CC oligosaccharides are useful as vaccines, therapeutics and diagnostic
 CC tools. Cells producing complex sialylated glycoproteins are useful for
 CC enhancing the value of heterologous expression systems and increasing the
 CC application of heterologous cell expression products as vaccines.
 CC Therapeutic and diagnostic tools as well as increasing the variety of
 CC heterologous proteins that can be produced and lowering biotechnology
 CC production costs. The present sequence represents a human sialic acid
 CC synthetase, which is used in the method of the invention
 XX
 SO Sequence 359 AA;
 XX
 Query Match 100.0%; Score 1887; DB 4; Length 359;
 Best Local Similarity 100.0%; Pred. No. 9,4e-184;
 Matches 359; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 1 MFLELELCPRWVGQHPFCFIIAIGQNHQGLDVAKRMIRKMEGACADCAKFKQSELEF 60
 |||||||

DB 1 MFLELELCPRWVGQHPFCFIIAIGQNHQGLDVAKRMIRKMEGACADCAKFKQSELEF 60
 QY 61 KENRKALRPTYSKHSAGKTYGENRKHLEPESHDQYRELQRYABEYGIPTTASGMDENAVE 120
 DB 61 KENRKALRPTYSKHSAGKTYGENRKHLEPESHDQYRELQRYABEYGIPTTASGMDENAVE 120
 QY 121 FLHEINVPFPKVGSGDTNNPFLYEKTAKGPRPMVTSQMGQMDTKQYQIVKPNPNC 180
 DB 121 FLHEINVPFPKVGSGDTNNPFLYEKTAKGPRPMVTSQMGQMDTKQYQIVKPNPNC 180
 QY 161 FLOCTSAVPLQEDYNLAVISEYOKLPDIPFGSGHETGIAISVAVALGAKVLERHIT 240
 DB 161 FLOCTSAVPLQEDYNLAVISEYOKLPDIPFGSGHETGIAISVAVALGAKVLERHIT 240
 QY 241 LDKTWKGSDBSASLPEGLAELVRSVLVERALGSPYQOLLPCENACNEKLGKSVYAKVK 300
 DB 241 LDKTWKGSDBSASLPEGLAELVRSVLVERALGSPYQOLLPCENACNEKLGKSVYAKVK 300
 QY 301 IPEGTILMDMLTYKVEBPXAYPPEDIFNLVGGKVLVYEBDDTMBELVDNHGKKITS 359
 DB 301 IPEGTILMDMLTYKVEBPXAYPPEDIFNLVGGKVLVYEBDDTMBELVDNHGKKITS 359
 RESULT 4
 AAO26546
 ID AAO26546 standard; protein; 359 AA.
 XX
 AC AAO26546;
 XX
 DT 06-MAR-2003 (first entry)
 XX
 DE Human SA synthetase protein.
 XX
 KM Vaccine; glycoprotein; insect cell; enzyme; N-acetylglucosamine-2;
 KM GlcNAc-2; epimerase; UDP-GlcNAc; mannosyl (Man)NAc; sialic acid;
 KM synthetase; aldolase; cytidine monophosphate-sialic acid; CMP-SA;
 KM transporter; sialylated glycoprotein; human.
 XX
 OS Homo sapiens.
 XX
 PN US2002142386-A1.
 PD 03-OCT-2002.
 XX
 PF 16-AUG-2001; 2001US-00930440.
 XX
 PR 02-MAR-1999; 99US-01225892P.
 PR 08-DEC-1999; 99US-0169624P.
 PR 25-AUG-2000; 2000US-0227579P.
 XX
 PA (BETE/) BETENBAUGH M J.
 PA (LAWR/) LAWRENCE S.
 PA (LEBY/) LEE Y C.
 PA (COLE/) COLEMAN T A.
 XX
 PI Betenbaugh MJ, Lawrence S, Lee YC, Coleman TA;
 PI WPI; 2003-102519/09.
 DR N-PSDB; AAL53993.
 XX
 PT Manipulating glycoprotein production in insect cell, involves enhancing
 PT expression of enzymes involved in carbohydrate processing pathway such as
 PT N-acetylglucosamine-2 epimerase or sialic acid synthetase.
 XX
 PS Claim 16; Fig 32; 88pp; English.
 XX
 CC The invention relates to a novel method for manipulating glycoprotein
 CC production in an insect cell comprising enhancing expression of an
 CC enzyme, such as N-acetylglucosamine-2 (GlcNAc-2) epimerase, one
 CC catalysing conversion of UDP-GlcNAc to mannose (Man)NAc, sialic acid
 CC synthetase, aldolase, cytidine monophosphate-sialic acid (CMP-SA)
 CC synthetase or CMP-SA transporter, where the expression of each enzyme is
 CC enhanced to above endogenous levels. The novel method is useful for

manipulating glycoprotein production in an insect cell. Further methods of the invention are useful for producing sialylated glycoprotein. The sialylated glycoprotein produced by the above mentioned methods are useful as pharmaceutical compositions, vaccines, diagnostics and therapeutics. This sequence represents the human SA synthetase protein of the invention

Sequence 359 AA:

Query Match 100.0%; Score 1887; DB 6; Length 359;
Best Local Similarity 100.0%; Pred. No. 9,4e-184;
Matches 359; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MPELELCGRWVGQHPCEPIIAEIGNHGDLDVAKRMIMAKKCGADCAKFOKSELEF 60
1 MPELELCGRWVGQHPCEPIIAEIGNHGDLDVAKRMIMAKKCGADCAKFOKSELEF 60
61 KENRKALERPPTSKHSGKTYGHEHKLHFFSHDOYRELQRYAEVGIFFTAGMDMAVE 120
61 KENRKALERPPTSKHSGKTYGHEHKLHFFSHDOYRELQRYAEVGIFFTAGMDMAVE 120
121 FLHELNVPPFKVSGSDTNPPYLKTAAGKRPMTISSGQSMOTKQVQIVKPLNPNFC 180
121 FLHELNVPPFKVSGSDTNPPYLKTAAGKRPMTISSGQSMOTKQVQIVKPLNPNFC 180
181 FLQCTSAVPLQPEBDVNLRVISEYOKLPDIPIGSGHETGIAISVAVALGAKVLERHIT 240
181 FLQCTSAVPLQPEBDVNLRVISEYOKLPDIPIGSGHETGIAISVAVALGAKVLERHIT 240
241 LDKTWKSDHSASLEPGEIAELVRSVLYERALGSPTRKQLPCENACNEKLGKSVAAVK 300
241 LDKTWKSDHSASLEPGEIAELVRSVLYERALGSPTRKQLPCENACNEKLGKSVAAVK 300
301 IPEGITLTMDMLTVKVGEPKAYPPEDIENLVGKKVLTVEEDDTIMEELVDNHGKKIKS 359
301 IPEGITLTMDMLTVKVGEPKAYPPEDIENLVGKKVLTVEEDDTIMEELVDNHGKKIKS 359

RESULT 5

ABR90221 ID ABR90221 standard; protein; 359 AA.

ABR90221; 24-MAY-2002 (first entry)

Human polypeptide SEQ ID NO 2597.

Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral; antitubercular; hepatotropic; antidiabetic; antinflammatory; antitumor; vulvar; anticonvulsant; antibacterial; antifungal; antiparasitic; cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder; neurological disease; infection; human; secreted protein.

Homo sapiens.

MO200190304-A2.

29-NOV-2001.

18-MAY-2001; 2001MO-US016450.

19-MAY-2000; 2000US-02055152.

(HUMA-) HUMAN GENOME SCI INC.

Birze CE, Rosen CA;

WPI; 2002-122018/16.

N-PSDB; ABL90630.

Novel 1405 isolated polypeptides, useful for diagnosis, treatment and prevention of neural, immune system, muscular, reproductive,

gastrointestinal, pulmonary, cardiovascular, renal and proliferative disorders.

Claim 11; SEQ ID NO 2597; 2081bp + Sequence Listing; English.

The invention relates to novel genes (ABL89449-ABL90853) and proteins (ABL89040-ABL90444) useful for preventing, treating or ameliorating medical conditions e.g. by protein or gene therapy. The genes are isolated from a range of human tissues disclosed in the specification. The nucleic acids, proteins, antibodies and (ant)agonists are useful in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and ovarian cancer and other cancers of the adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune disorders e.g. Addison's disease, allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c) cardiovascular disorders such as myocardial ischaemias; (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f) infectious diseases such as viral, bacterial, fungal and parasitic infections. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WPI at http://wipo.int/pub/published_pct_sequences

Sequence 359 AA;

Query Match 99.8%; Score 1883; DB 5; Length 359;
Best Local Similarity 99.7%; Pred. No. 2,4e-183;
Matches 358; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 MPELELCGRWVGQHPCEPIIAEIGNHGDLDVAKRMIMAKKCGADCAKFOKSELEF 60
1 MPELELCGRWVGQHPCEPIIAEIGNHGDLDVAKRMIMAKKCGADCAKFOKSELEF 60
61 KENRKALERPPTSKHSGKTYGHEHKLHFFSHDOYRELQRYAEVGIFFTAGMDMAVE 120
61 KENRKALERPPTSKHSGKTYGHEHKLHFFSHDOYRELQRYAEVGIFFTAGMDMAVE 120
121 FLHELNVPPFKVSGSDTNPPYLKTAAGKRPMTISSGQSMOTKQVQIVKPLNPNFC 180
121 FLHELNVPPFKVSGSDTNPPYLKTAAGKRPMTISSGQSMOTKQVQIVKPLNPNFC 180
181 FLQCTSAVPLQPEBDVNLRVISEYOKLPDIPIGSGHETGIAISVAVALGAKVLERHIT 240
181 FLQCTSAVPLQPEBDVNLRVISEYOKLPDIPIGSGHETGIAISVAVALGAKVLERHIT 240
241 LDKTWKSDHSASLEPGEIAELVRSVLYERALGSPTRKQLPCENACNEKLGKSVAAVK 300
241 LDKTWKSDHSASLEPGEIAELVRSVLYERALGSPTRKQLPCENACNEKLGKSVAAVK 300
301 IPEGITLTMDMLTVKVGEPKAYPPEDIENLVGKKVLTVEEDDTIMEELVDNHGKKIKS 359
301 IPEGITLTMDMLTVKVGEPKAYPPEDIENLVGKKVLTVEEDDTIMEELVDNHGKKIKS 359

RESULT 6

AAM39986 ID AAM39986 standard; protein; 359 AA.

AAM39986; 22-OCT-2001 (first entry)

Human polypeptide SEQ ID NO 3131.

Human; nootropic; immunosuppressant; cyostatic; gene therapy; cancer; peripheral nervous system; neuropathy; central nervous system; CNS; Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic; amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic; chemokinetic; thrombolytic; drug screening; arthritis; inflammation; leukaemia.

Homo sapiens.

PN WO200153312-A1.
 XX
 PD 26-JUL-2001.
 XX
 PF 26-DEC-2000; 2000MO-US034263.
 XX
 PR 23-DEC-1999; 99US-00471275.
 PR 21-JAN-2000; 2000US-00488725.
 PR 25-APR-2000; 2000US-00552317.
 PR 20-JUN-2000; 2000US-00586042.
 PR 19-JUL-2000; 2000US-00620312.
 PR 03-AUG-2000; 2000US-00653450.
 PR 14-SEP-2000; 2000US-00662191.
 PR 19-OCT-2000; 2000US-00693036.
 PR 29-NOV-2000; 2000US-00727344.
 XX
 PA (HSE-) HSEQ INC.
 XX
 PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
 PI Zhou P, Goodrich R, Dmanac RT;
 DR WPI; 2001-442253/47.
 XX
 DR N-PSDB; AA159142.
 XX
 PT Novel nucleic acids and polypeptides, useful for treating disorders such
 PT as central nervous system injuries.
 XX
 PS Example 4; SEQ ID NO 3131; 10078bp; English.
 XX
 CC The invention relates to human nucleic acids (AA157798-AA161369) and the
 CC encoded polypeptides (AA038642-AA042233) with neurotrophic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localized neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Activating/inhibiting activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening and
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders. Note: The sequence data for this patent did not form
 CC part of the printed specification
 CC
 XX
 SQ Sequence 359 AA;
 Query Match 99.4%; Score 1880; DB 4; Length 359;
 Best Local Similarity 99.4%; Pred. No. 4.9e-183; Indels 0; Gaps 0;
 Matches 357; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MPELELCPCRWYGGCHPCPIIAEIQNHQGDLDVAKMTIMAKCEGACDCAKQKSELEF 60
 DB 1 MPELELCPCRWYGGCHPCPIIAEIQNHQGDLDVAKMTIMAKCEGACDCAKQKSELEF 60
 QY 61 KPRKRLDRYTSKSHGKTYGKHRLSHSHOYBELGYAREVGFPTASGMDMAVE 120
 DB 61 KPRKRLDRYTSKSHGKTYGKHRLSHSHOYBELGYAREVGFPTASGMDMAVE 120
 QY 61 KPRKRLDRYTSKSHGKTYGKHRLSHSHOYBELGYAREVGFPTASGMDMAVE 120
 DB 61 KPRKRLDRYTSKSHGKTYGKHRLSHSHOYBELGYAREVGFPTASGMDMAVE 120
 QY 121 FLHLELVNPFKVGSGDTNNPFLYEKTAKKRPMVVISGMSQMDTMKQVQIVKPLNPNFC 180
 DB 121 FLHLELVNPFKVGSGDTNNPFLYEKTAKKRPMVVISGMSQMDTMKQVQIVKPLNPNFC 180
 QY 121 FLHLELVNPFKVGSGDTNNPFLYEKTAKKRPMVVISGMSQMDTMKQVQIVKPLNPNFC 180
 DB 121 FLHLELVNPFKVGSGDTNNPFLYEKTAKKRPMVVISGMSQMDTMKQVQIVKPLNPNFC 180
 QY 181 FLOCTSNAPYQEDVNLKRVSEYOKLFPDIPITVSGHETGIALSVAAVLAAGAVLERHIT 240
 DB 181 FLOCTSNAPYQEDVNLKRVSEYOKLFPDIPITVSGHETGIALSVAAVLAAGAVLERHIT 240
 QY 181 FLOCTSNAPYQEDVNLKRVSEYOKLFPDIPITVSGHETGIALSVAAVLAAGAVLERHIT 240
 DB 181 FLOCTSNAPYQEDVNLKRVSEYOKLFPDIPITVSGHETGIALSVAAVLAAGAVLERHIT 240
 QY 241 LDKTWKSDHSASLEPELAEIVRSVLRALGSPTKQLLPCEAENKRGKSVYAKYK 300
 DB 241 LDKTWKSDHSASLEPELAEIVRSVLRALGSPTKQLLPCEAENKRGKSVYAKYK 300
 QY 301 IPEGTILTMMLTVKGEPRKYPEDIFNLVGKVLVTVEEDTILMEELVNHGKIKS 359
 DB 301 IPEGTILTMMLTVKGEPRKYPEDIFNLVGKVLVTVEEDTILMEELVNHGKIKS 359

DB 301 IPEGTILTMMLTVKGEPRKYPEDIFNLVGKVLVTVEEDTILMEELVNHGKIKS 359
 RESULT 7
 ID AAB93183 standard; protein; 359 AA.
 XX
 AC AAB93183;
 XX
 DT 26-JUN-2001 (first entry)
 XX
 DE Human protein sequence SEQ ID NO:12130.
 XX
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
 XX
 OS Homo sapiens.
 XX
 PN EPI074617-A2.
 PD 07-FEB-2001.
 XX
 PF 28-JUL-2000; 2000BP-00116126.
 XX
 PR 29-JUL-1999; 99BP-00248036.
 PR 27-AUG-1999; 99BP-00300253.
 PR 11-JAN-2000; 2000JP-00118776.
 PR 02-MAY-2000; 2000JP-00183767.
 PR 09-JUN-2000; 2000JP-00241899.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX
 DR WPI; 2001-118749/34.
 XX
 PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-
 PT length cDNAs defined in the specification, and for the detection and/or
 PT diagnosis of the abnormality of the proteins encoded by the full-length
 PT cDNAs.
 XX
 PS Claim 8; SEQ ID NO 12130; 2537bp + Sequence listing; English.
 XX
 CC The present invention describes primer sets for synthesizing 5602 full-
 CC length cDNAs defined in the specification. Where a primer set comprises:
 CC (a) an oligo-dT primer and an oligonucleotide complementary to the
 CC complementary strand of a polynucleotide which comprises one of the 5602
 CC nucleotide sequences defined in the specification; where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in the
 CC specification. The primer sets can be used in antisense therapy and in
 CC gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
 CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
 CC oligonucleotides, all of which are used in the exemplification of the
 CC present invention
 CC
 XX
 SQ Sequence 359 AA;
 Query Match 99.4%; Score 1876; DB 4; Length 359;
 Best Local Similarity 99.2%; Pred. No. 1.3e-182; Indels 0; Gaps 0;
 Matches 356; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

XX	Query Match	94.8%;	Score 1789;	DB 4;	Length 400;
CC	Best Local Similarity	88.3%;	Pred. No. 1.2e-173;		
CC	Matches 346;	Conservative 0;	Mismatches 2;	Indels 44;	Gaps 1;
XX	The invention relates to human nucleic acids (AA157798-AA161369) and the				
CC	encoded polypeptides (AAM8642-AAM4213) with neurotropic				
CC	immunosuppressant and cytostatic activity. The polynucleotides are useful				
CC	in gene therapy. A composition containing a polypeptide or polynucleotide				
CC	of the invention may be used to treat diseases of the peripheral nervous				
CC	system, such as peripheral nervous injuries, peripheral neuropathy and				
CC	localised neuropathies and central nervous system diseases, such as				
CC	Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic				
CC	lateral sclerosis, and Shy-Drager Syndrome. Other uses include the				
CC	utilisation of the activities such as: Immune system suppression,				
CC	Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic				
CC	and thrombolytic activity, cancer diagnosis and therapy, drug screening,				
CC	assays for receptor activity, arthritis and inflammation, leukaemia, and				
CC	C.N.S disorders. Note: The sequence data for this patent did not form				
CC	part of the printed specification				
XX	Sequence 400 AA:				
QY	1	MPLEIELCPGHWVGCGHPCPIIABIIGNHGDDLVAKRMIRMAKCGADCKPKXSLRF	60		
DB	9	MPLEIELCPGHWVGCGHPCPIIABIIGNHGDDLVAKRMIRMAKCGADCKPKXSLRF	68		
QY	61	KENKALERPYTSKXSMCKTYGKHKRLFEHQYRZLQRYAEVGGIFFTASGMDENAVE	120		
DB	69	KENKALERPYTSKXSMCKTYGKHKRLFEHQYRZLQRYAEVGGIFFTASGMDENAVE	128		
QY	121	FLHEINVPFFKVGSGDTNNPEYLEKTKX-----	148		
DB	129	FLHEINVPFFKVGSGDTNNPEYLEKTKXGMSVLRDVCVQUNDETSMDVLGRVTS	188		
QY	149	-----KRPWVTSKQMSMDPKQVOIVKPLNPFCELOCTSAIYPLQPEDVN	196		
DB	189	KERKLVAVLDYSGEPWVTSKQMSMDPKQVOIVKPLNPFCELOCTSAIYPLQPEDVN	248		
QY	197	LRVISEYQKLPDDIPIGVSGHETGIAISVAVALGAKVLEHHITLDTKWSGDSASLEP	256		
DB	249	LRVISEYQKLPDDIPIGVSGHETGIAISVAVALGAKVLEHHITLDTKWSGDSASLEP	308		
QY	257	GELAEIVRSVAVRRALASPTKQILLPEMAENELGGSVAKYKIPSTILITMMLTVKV	316		
DB	309	GELAEIVRSVAVRRALASPTKQILLPEMAENELGGSVAKYKIPSTILITMMLTVKV	368		
QY	317	GEPRAYPPEDIEFNLVGKKVLTVEEDDTIMEE	348		
DB	369	GEPRAYPPEDIEFNLVGKKVLTVEEDDTIMEE	400		
XX	RESULT 9				
XX	ABB71364	ID	ABB71364	standard; protein; 338 AA.	
XX	ABB71364;				
XX	26-MAR-2002	(first entry)			
XX	Drosophila melanogaster polypeptide SEQ ID NO 40884.				
XX	Drosophila; developmental biology; cell signalling; insecticide;				
XX	pharmaceutical.				
XX	Drosophila melanogaster.				
XX	WO200171042-A2.				
XX	27-SEP-2001.				
XX	23-MAR-2001; 2001IMO-US009231.				

PR 23-MAR-2000; 2000US-0191637P.
 PR 11-JUL-2000; 2000US-00614150.
 XX
 XX (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li FWD, Myers EM;
 XX
 DR WPI; 2001-656860/75.
 DR N-PSDB; ABL15467.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signaling and cell-cell
 PT interactions.
 XX
 PS Disclosure; SEQ ID NO 40884; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signaling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB116176-AB120511), expressed DNA
 CC sequences (AB101840-AB116175) and the encoded proteins (AB557737-
 CC AB872072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 CC
 XX
 XX Sequence 338 AA;
 SQ
 Query Match 41.5%; Score 783; DB 4; Length 338;
 Best Local Similarity 45.6%; Pred. No. 6,5e-71;
 Matches 154; Conservative 60; Mismatches 100; Indels 24; Gaps 3;
 QY 39 MIMAKECGADCAKQSELEPFKRNALBPYTSKSKGKTEGKHLPSHDQREL 98
 DB 1 MIMAKKAGCHCVFQKSDLPKFTPSALDREYISDIAWKGTYGHEHVEFSNDQYLQ 60
 QY 99 QYVAEYGFPTASGMEVAEVEFLHNVPEFKYSGDINNFPLEKTKKGRPMVTS 158
 DB 61 QAKKEINVDFTANMDEBSIEFLSALNVEPIKISGDANNFLKAAALNLEPLVISTG 120
 QY 159 MOSMDTKQVQYQVKEP-NENFCLOCTSAVPILOPEEDNLRVISEYKLEPPDIPIGYSGH 217
 DB 121 MQMTQVTERIVQYTRRESKEXDYALMHCVSYPTDPKDCSLQILISVLRFPNVIAIGYSGH 180
 QY 218 ETGALISVVAALADAKYLBHHTLDKTKWGSDBASISPGELALVNSV----- 266
 DB 181 ELGVITISQAAVLLGARIVERHFTLDKSKQSDHSCSELPQELKALTALITNFKLSVMP 240
 QY 267 -RIVERALGS-----PTKQLPCENACNEKIKSVAVAKVKEGTILTMDELTV 314
 DB 241 PQELVYKNGDEBHEALQHVESKTIIPCELPCKNKGKSIIVAAARNINKGVRLQADMAI 300
 QY 315 KVGEPKAVPEPDITNLYGKQVLVVEEDDTIMEELVDN 352
 DB 301 KVSEPSGILTAEDPDLVGKELADNIGDEPILGNSIIN 338
 DB
 RESULT 10
 ID AAY68965 standard; protein, 338 AA.
 AC AAY68965;
 XX
 XX 30-MAY-2000 (first entry)
 DE Cps2P protein which is involved in sialic acid synthesis.
 XX
 XX Capsular gene cluster; serotype 2; polysaccharide biosynthesis;
 XX capsular component; antigen; regulation; chain length determination;
 XX complement-mediated opsonophagocytosis; serotype-specific detection;
 XX antigen; vaccine; Streptococcal disease; ORF 22; ORF 2Y; ORF 2Z; Cps2A;
 KW Cps2B; Cps2C; Cps2D; Cps2E; Cps2F; Cps2G; Cps2H; Cps2I; Cps2J; Cps2K;

KM Cps2O; Cps2P; Cps2Q; Cps2R; Cps2S; Cps2T.
 XX Streptococcus suis.
 OS
 XX MO200005378-A2.
 XX
 XX 03-FEB-2000.
 XX
 XX 19-JUL-1999; 99MO-NL000460.
 XX
 XX 22-JUL-1998; 98EP-00202465.
 PR 22-JUL-1998; 98EP-00202467.
 XX
 XX (DIEN-) STICHTING DIENST LANDBOUDKUNDIG ONDERZOE.
 PA Smith HE;
 XX
 PI WPI; 2000-195104/17.
 DR N-PSDB; AA260929.
 XX
 XX New nucleic acid containing the capsular gene cluster of Streptococcus
 PT suis, used for serotype-specific detection and to generate antigens or
 PT mutants for vaccination.
 XX
 XX Disclosure; Fig 3; 14pp; English.
 PS
 XX The proteins AAY68950-69 are encoded by the capsular gene cluster of
 CC Streptococcus suis serotype 2. The genes in this cluster are involved in
 CC polysaccharide biosynthesis of capsular components and antigens. The
 CC proteins are involved in regulation (CpsA), chain length determination
 CC (CpsB, CpsC), export (CpsD), and biosynthesis (CpsE, CpsF, CpsG, CpsH,
 CC CpsJ, CpsK). The capsule confers bacterium resistance to complement-
 CC mediated opsonophagocytosis. The gene cluster is used as a source of
 CC probes and primers for serotype-specific detection of S. suis and is also
 CC useful for recombinant production of the proteins. The proteins are then
 CC useful for producing antigens that can be used in vaccines, for
 CC controlling or eradicating a Streptococcal disease, in humans or animals,
 CC e.g. against S. suis in pigs
 CC
 XX
 XX Sequence 338 AA;
 SQ
 Query Match 28.9%; Score 544.5; DB 3; Length 338;
 Best Local Similarity 36.0%; Pred. No. 1.6e-46;
 Matches 123; Conservative 73; Mismatches 129; Indels 17; Gaps 8;
 QY 20 FTIABIGONHGDLDYAKMIRAKCGADCAKQSELEPFKRNALBPYTSKSKGK 79
 DB 3 YTIABIGONHGDYHAKKRVAVAVDCGDVAFQFFKADLISKAPAEY-QKITGE 61
 QY 80 TYG-EHKHGLEPSHDQYRELQRYAEVGIFFTASGMDENAEVFLHNLVPEFKVSGDT 137
 DB 62 SDQLEWTRLELSPREYLDLRDYCLEKGVDSYTFEEDSLDFLSTMPVYKISGEL 121
 QY 138 NNEPVEKTRAKGRPMVVISGMSQMDTMKOVQIVPLINDN---PCFLOCTSAVPILOPE 193
 DB 122 TKLPYEKGRQKKKYLSTGMAVMD---RHHAVAILQENGTDTISILCTEYERPP 178
 QY 194 DVNLRVISEYQKLPDIPIGYSGHGTGIALISVAVALGAKVERHHTLDTKTKGSDHAS 253
 DB 179 ALNLNVLHTLKEFPNLTIGSDHSVSEVPVIAAAMGAEILIKHFTLDNEMGPDHKAS 238
 QY 254 LEPGELAEIVASRYVERALGSPTKQLPCENACNEKLT--GKSVAKVKEGTILTM 311
 DB 239 AFPDITLALVKGRIYVQSLKTEKE--DEEVARKYVAKRSIVAKKALAKGEVTFEEN 296
 QY 312 LTVKVGSP-KAYPPEDIFFNLVGKQVLVVEEDDTIMEELVDN 352
 DB 297 ITVK--RPGKGISPMWEYKVLGVSEODPREDQNI CHSAFEN 336
 DB
 RESULT 11
 ID ABP26810 standard; protein, 341 AA.
 ABP26810

XX ABP26810;
AC
XX
XX 02-JUL-2002 (first entry)
DT
XX
XX Streptococcus polypeptide SEQ ID NO 2796.
DE
XX Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
XX group A streptococcus; Streptococcus pyogenes; antibacterial;
XX anti-inflammatory; infection; vaccine; meningitis; gene therapy.
XX
XX Streptococcus agalactiae.
OS
XX WO200234771-A2.
FN
XX
XX 02-MAY-2002.
PD
XX
XX 29-OCT-2001; 2001WO-GB004789.
PF
XX
XX 27-OCT-2000; 2000GB-00026333.
PR
XX 24-NOV-2000; 2000GB-00028727.
PR 07-MAR-2001; 2001GB-00005640.
XX
XX (CHIR-) CHIRON SPA.
XX (GENO-) INST GENOMIC RES.
XX
XX Telford J, Maignant V, Margatit Y Rosl, Grandi G, Fraser C;
PI Telford J;
XX
XX WPI; 2002-352536/38.
XX N-PSDB; ABN67441.
DR
XX
XX New Streptococcus protein for the treatment or prevention of infection or
PT disease caused by Streptococcus bacteria, such as meningitis, and for
PT detecting a compound that binds to the protein.
XX
XX Claim 1; Page 3430; 4525BP; English.
XX
XX The invention relates to a protein (ABP25413-ABP30895) from group B
XX streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
XX (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
XX the specification. The proteins have antibacterial and anti-inflammatory
XX activity. (I), nucleic acids encoding (II), ABN6044-ABN71526 and
XX antibodies that bind (I) are used in the manufacture of medicaments for
XX the treatment or prevention of infection or disease caused by
XX Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
XX Streptococcus agalactiae and S. pyogenes.
XX Nucleic acids encoding (I) are used to detect Streptococcus in a
XX biological sample. (I) is used to determine whether a compound binds to
XX (I). A composition comprising (I) or a nucleic acid encoding (I), may be
XX used as a vaccine or diagnostic composition. The disease caused by
XX Streptococcus that is prevented or treated may be meningitis. Nucleic
XX acid encoding (I) may be used to recombinantly produce (I) and may be
XX used in gene therapy. Antibodies to (I) are used for affinity
XX chromatography, immunoassays, and distinguishing/identifying
XX Streptococcus proteins
XX
XX Sequence 341 AA;
SQ

Query Match 28.2%; Score 532.5; DB 5; length 341;
Best Local Similarity 36.8%; Pred. No. 2.7e-45;
Matches 125; Conservative 66; Mismatches 136; Indels 13; Gaps 6;

DB 3 YIIAIGCHNHGDIINAKKRVDAVSCGVDAVVFQTFKAKISKFAPAEY-QRAITGT 61
QY 20 FIIAIGCHNHGDIINAKKRVDAVSCGVDAVVFQTFKAKISKFAPAEY-QRAITGT 79
DB 80 TVG--EAKHLESHDQYRIARAEVGIFFTAAGDMMAVEFLHNLVPPFKVSGSDT 137
QY 62 ASQGLMTBRLBLSFEYILMDRYALSKVETSTFDESESLFISTMPYKIPSGRI 121
DB 138 NNPFLYKAKKRPVWISSGQSMPTMGOVQIVKGLPN---PCFQICSAVPLQPE 193
QY 122 TNPFLYKAKKRPVWISSGQSMPTMGOVQIVKGLPN---PCFQICSAVPLQPE 193

QY 194 DVNLRVISEYQKLPDIPIDIGSGHETGIAISVAALGAKTLERHITLDTKWSGDSAS 253
DB 179 SLINLVHTLKDREKDLITIGSDHSISEVPIAAAGAEVIEHFTLDTMGEQDHAS 238
QY 254 LEPGELAEVRSVLYVERALGSPFKQILPCMACNENKIGKSVKVKIPESTILTMMLT 313
DB 239 ATPDILALVKGAVIVGQALGRPEKIPDVEERKKIVARSSVVALPKKGDIVSIENIT 298
QY 314 VKGEP-KAIPPEIDFNLVKGKVLVYBEDDTIMEELVDN 352
DB 239 VK--RPGNGISPMWYDILGQEAQDDFEEDVIRDSRFEN 336

RESULT 12
ID AA96102 standard; protein, 346 AA.
AA96102;
AC
XX
XX 19-DEC-2000 (first entry)
DT
XX
XX Escherichia coli sialic acid synthetase.
DE
XX Sialic acid synthetase; human; sas gene; sialylation; glycoprotein;
XX plasmidogen; transferrin; thyrotropin; Na⁺, K⁺-ATPase.
XX
XX Escherichia coli.
OS
XX WO200052135-A2.
FN
XX
XX 08-SEP-2000.
PD
XX
XX 01-MAR-2000; 2000WO-US005313.
PF
XX
XX 02-MAR-1999; 99US-012582P.
PR 08-DEC-1999; 99US-0169624P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX (UYJO-) UNIT JOHNS HOPKINS.
XX (UYWY-) UNIT WYOMING.
XX
XX Betenbough MJ, Lawrence S, Lee YC, Jarvis D, Coleman TA;
PI WPI; 2000-572178/53.
DR N-PSDB; AAA50570.
XX
XX Recombinant production of sialylated glycoproteins using cells in which
PT the expression of enzymes, e.g. sialic acid synthetase, involved in the
PT sialylation reaction has been altered.
XX
XX Example 6; Page 108; 144pp; English.
XX
XX The present sequence is that of Escherichia coli sialic acid synthetase,
XX encoded by the nub gene (see AAA50570). The nub gene was used to
XX isolate the corresponding human gene (see AAA50569) for sialic acid
XX synthetase (see AA96101). The invention provides methods and
XX recombinantly engineered cells for producing glycoproteins having
XX sialylated oligosaccharides. The methods involve altering the expression
XX of enzymes involved in carbohydrate processing. A claimed cell producing
XX sialylated glycoprotein at above endogenous levels expresses at least 1
XX (preferably human) enzyme selected from GlcNAc-2 epimerase, an enzyme
XX catalyzing the conversion of UDP-GlcNAc to ManNAc, sialic acid
XX synthetase, aldolase, CMP-sialic acid synthetase and CMP-sialic acid
XX transporter at above endogenous levels. A claimed method for manipulating
XX glycoprotein in an insect cell comprises enhancing the expression of 1 of
XX the above enzymes, and a claimed method for producing sialylated
XX glycoproteins involves expressing a heterologous protein (especially
XX plasmidogen, transferrin, Na⁺, K⁺-ATPase or thyrotropin) in the insect
XX cell. Yeast, insect, fungal, plant and bacterial host cells can be
XX engineered to produce new forms of sialylated glycoproteins, higher
XX concentrations of sialylated glycoproteins and/or elevated concentrations
XX of donor substrates (e.g. nucleotide sugars) required for sialylation

XX Sequence 346 AA;
 Query Match 27.1%; Score 510.5; DB 3; Length 346;
 Best Local Similarity 36.3%; Pred. No. 4.9e-43;
 Matches 123; Conservative 61; Mismatches 140; Indels 15; Gaps 6;

20 FIIAEIGQNHQDLDVAKKIMPAKCEGADCAKFOKSELEFFNRKALERPYSKISWG- 78
 5 YIIAEIGCNHNSVDIAREMILKAKAGVNAVVFQTFKADKLISALAPAEYQIKATGEL 64
 79 KTYGKRLHLEFSDQYRELQRYAEVGIPTFASGMDENAVFLHLELNPFFVSGDITN 138
 65 ESQLEMTKRLKEMKYDYILHMEYAVSLNLDVFSTPDEDSIDFLASIKOKIMTI PGCELL 124
 139 NFPYLEKTAKGRP---WVISGQMSMDTMKQYQIV---KPLNPFCLQCTSAVPLQ 192
 125 NLPYLEKIALPLPDKKIITSTGMATIDEIKOSVSTFNKKVPVGNITILHCTEPTFP 184
 193 EDVNLRYISEYOKLPDIPPIGSGHETGALISVAVALGAKVLERHITLDTKWKSGDHA 252
 185 EDVNLAINDLKGFPPKNIIGFSDHSGFYAALAAVPGYITFEKHFLLDKMSGPDHAA 244
 253 SLRPGELAEVRSVRLVERALGSPTKQLPCEMACNEKLGKSVAVAKIPEGITLMDYL 312
 245 SIPEDELKELCTIGRCVEKSLGNSKVVTSERKNKIIVARKSIITAKTEIKGSEVSEKNI 304
 313 TVKVGEP-KAYPPEDI FNLVGGKVLVTEEDDTIMEELV 350
 305 TTK--RPGNGISPMEMWNLGK-----IAEQDFIPDELI 336

RESULT 13
 AAB84684
 ID AAB84684 standard; protein; 346 AA.
 AAB84684;
 17-SEP-2001 (first entry)

17-SEP-2001 (first entry)

Amino acid sequence of a bacterial sialic acid synthetase.

Carbohydrate processing; sialylation; sialylated glycoprotein; CMP-SA; cytidine monophosphate-sialic acid; N-acetylglucosamine-2 epimerase; sialic acid synthetase; aldolase; CMP-SA synthetase; CMP-SA transporter; vaccine.

Escherichia coli.
 WO200142492-A1.
 14-JUN-2001.
 07-DEC-2000; 2000WC-US033136.
 09-DEC-1999; 99US-0169839P.
 (HUMA-) HUMAN GENOME SCI INC.
 (UYCO) UNIV JOHNS HOPKINS.
 (UTEM) UNIV TEMPLE.
 (UYW-) UNIV WYOMING.
 Betenbaugh MJ, Lawrence S, Lee YC, Coleman TA, Falter K; Jarvis D;
 WPI, 2001-441575/47.
 N-PSDB; AAB28459.

Cells producing cytidine monophosphate-sialic acid and sialylated glycoprotein above endogenous levels for production of vaccines and therapeutics.

Example 5; Page 165-166; 182pp; English.

XX The specification describes a method for manipulating carbohydrate processing pathways in cells of interest. The methods are used to manipulate multiple pathways involved with the sialylation reaction by using recombinant DNA technology and substrate feeding approaches to enable the production of sialylated glycoproteins in the cells. The sialylation process involves the post-translational addition of the donor substrate cytidine monophosphate-sialic acid (CMP-SA) onto a specific acceptor carbohydrate. The cells express at least one enzyme, selected from N-acetylglucosamine-2 epimerase, sialic acid synthetase, aldolase, CMP-SA synthetase and CMP-SA transporter, above endogenous levels. The cells are useful for producing complex sialylated glycoproteins in cells of interest, especially insect cells. Glycoproteins containing sialylated oligosaccharides are useful as vaccines, therapeutics and diagnostic tools. Cells producing complex sialylated glycoproteins are useful for enhancing the value of heterologous expression systems and increasing the application of heterologous cell expression products as vaccines, therapeutics and diagnostic tools as well as increasing the variety of heterologous proteins that can be produced and lowering biotechnology production costs. The present sequence represents a sialic acid synthetase (neub), which is used in the method of the invention

SO Sequence 346 AA;
 Query Match 27.1%; Score 510.5; DB 4; Length 346;
 Best Local Similarity 36.3%; Pred. No. 4.9e-43;
 Matches 123; Conservative 61; Mismatches 140; Indels 15; Gaps 6;

20 FIIAEIGQNHQDLDVAKKIMPAKCEGADCAKFOKSELEFFNRKALERPYSKISWG- 78
 5 YIIAEIGCNHNSVDIAREMILKAKAGVNAVVFQTFKADKLISALAPAEYQIKATGEL 64
 79 KTYGKRLHLEFSDQYRELQRYAEVGIPTFASGMDENAVFLHLELNPFFVSGDITN 138
 65 ESQLEMTKRLKEMKYDYILHMEYAVSLNLDVFSTPDEDSIDFLASIKOKIMTI PGCELL 124
 139 NFPYLEKTAKGRP---WVISGQMSMDTMKQYQIV---KPLNPFCLQCTSAVPLQ 192
 125 NLPYLEKIALPLPDKKIITSTGMATIDEIKOSVSTFNKKVPVGNITILHCTEPTFP 184
 193 EDVNLRYISEYOKLPDIPPIGSGHETGALISVAVALGAKVLERHITLDTKWKSGDHA 252
 185 EDVNLAINDLKGFPPKNIIGFSDHSGFYAALAAVPGYITFEKHFLLDKMSGPDHAA 244
 253 SLRPGELAEVRSVRLVERALGSPTKQLPCEMACNEKLGKSVAVAKIPEGITLMDYL 312
 245 SIPEDELKELCTIGRCVEKSLGNSKVVTSERKNKIIVARKSIITAKTEIKGSEVSEKNI 304
 313 TVKVGEP-KAYPPEDI FNLVGGKVLVTEEDDTIMEELV 350
 305 TTK--RPGNGISPMEMWNLGK-----IAEQDFIPDELI 336

RESULT 14
 AAO26547
 ID AAO26547 standard; protein; 346 AA.
 AAO26547;
 06-MAR-2003 (first entry)

06-MAR-2003 (first entry)

Bacterial sialic acid synthetase (Neub) protein.

Vaccine; glycoprotein; insect cell; enzyme; N-acetylglucosamine-2; sialic acid; epimerase; UDP-GlcNAc; mannose; (Man)NAC; sialic acid; synthetase; aldolase; cytidine monophosphate-sialic acid; CMP-SA; transporter; sialylated glycoprotein.

Escherichia coli.
 US2002142386-A1.
 03-OCT-2002.

XX 16-AUG-2001; 2001US-00930440.
 PF 02-MAR-1999; 99US-0122582P.
 XX 08-DEC-1999; 99US-0169624P.
 PR 25-AUG-2000; 2000US-0227579P.
 XX (BRET/) BETENBAUGH M J.
 PA (LAMR/) LAWRENCE S.
 PA (LEBY/) LEE Y C.
 PA (COLE/) COLEMAN T A.
 XX Betenbaugh MJ, Lawrence S, Lee YC, Coleman TA;
 PI WPI: 2003-102519/09.
 DR N-PSDB; AALS394.
 XX Manipulating glycoprotein production in insect cell, involves enhancing
 PT expression of enzymes involved in carbohydrate processing pathway such as
 PT N-acetylglucosamine-2 epimerase or sialic acid synthetase.
 XX Disclosure; Fig 35D; 86pp; English.
 PS The invention relates to a novel method for manipulating glycoprotein
 CC production in an insect cell comprising enhancing expression of an
 CC enzyme, such as N-acetylglucosamine-2 (GlcNAc-2) epimerase, one
 CC catalysing conversion of UDP-GlcNAc to mannose (Man)Nac, sialic acid
 CC synthetase, aldolase, cytidine monophosphate-sialic acid (CMP-SA)
 CC synthetase or CMP-SA transporter, where the expression of each enzyme is
 CC enhanced to above endogenous levels. The novel method is useful for
 CC manipulating glycoprotein production in an insect cell. Further methods
 CC of the invention are useful for producing sialylated glycoprotein. The
 CC sialylated glycoprotein produced by the above mentioned methods are
 CC useful as pharmaceutical compositions, vaccines, diagnostics and
 CC therapeutics. This sequence represents the bacterial sialic acid (NeuB)
 CC synthetase protein of the invention
 XX
 XX Sequence 346 AA;
 SQ
 Query Match 27.1%; Score 510.5; DB 6; Length 346;
 Best Local Similarity 36.3%; Pred. No. 4.9e-43;
 Matches 123; Conservative 61; Mismatches 140; Indels 15; Gaps 6;
 QY 20 FTIATIGNHQGDLDVAKRMIMAKECGADCKAFKXSEIEFFKFKALERTPTSGHSG- 78
 DB 5 YIVARIQCHNGSVIARIMILKAEAGVNAVVFQTFKADKLISAIAPAEYQIINTSEL 64
 QY 79 KTYGEHKRLERSHQYRELQYAEVGIFFTASGMDENAVFLHELNVPPFKVSGSDTN 138
 DB 65 BQOLEMTKLEMKIDYILHLMYANSINDVSTFDESDIDFLASLQKIKWISGELL 124
 QY 139 NFPLYEKTAKGRP--NVISGQMSMDTKOVQIV--KPLNFNFCLOCTAIVYLP 192
 DB 125 NLPYLEKIKLPIPKKIIISTGMATIDIKQSVSIFINNKKVPVGNITILHNCITETPTPF 184
 QY 193 EDVNLKVISSEYQKLPFDIPIGVSGHETGIAISVAVALGAKYLERHTLTDKTKWGSDSA 252
 DB 185 EDVNLKVISSEYQKLPFDIPIGVSGHETGIAISVAVALGAKYLERHTLTDKTKWGSDSA 244
 QY 253 SLEPGEALVRSVTLVERALQSPKQILLPCMAENCKLQKGVAKVPIEGTILTMML 312
 DB 245 SLEPGEALVRSVTLVERALQSPKQILLPCMAENCKLQKGVAKVPIEGTILTMML 304
 QY 313 TVKVGEP-KAVPREDIPNLVKKVLTVEEDDTIMEILV 350
 DB 305 TTK--RPGNGISPEKMYNLGK-----IAQDFIDBELI 336

RESULT 15
 ABG17251
 XX ABG17251 standard; protein; 125 AA.
 AC ABG17251;

XX 16-FEB-2002 (first entry)
 DT Novel human diagnostic protein #17242.
 DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX Homo sapiens.
 XX WO200175067-A2.
 EN 11-OCT-2001.
 PD 30-MAR-2001; 2001WO-US008631.
 PF 31-MAR-2000; 2000US-00540217.
 PR 23-AUG-2000; 2000US-00649167.
 XX (HYSE-) HYSEQ INC.
 XX Dmanac RT, Liu C, Tang YT;
 PI WPI: 2001-639362/73.
 DR N-PSDB; AALS1438.
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX Claim 20; SEQ ID NO 47610; 103pp; English.
 PS The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptides and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG030377 represent novel human diagnostic
 CC amino acid sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pat_sequences
 XX
 XX Sequence 125 AA;
 SQ
 Query Match 22.2%; Score 419.5; DB 4; Length 125;
 Best Local Similarity 79.2%; Pred. No. 2e-34;
 Matches 80; Conservative 6; Mismatches 14; Indels 1; Gaps 1;
 QY 1 MPELELCPPGHWGQHPCCFIARIGNHQGDLDVAKRMIMAKECGADCKAFKXSEIEF 60
 DB 9 MPVELELCPGHWGQHPCCFIARIGNHQGDLDVAKRMIMAKECGADCKAFKXSEIEF 68
 QY 61 KPNKALRPYTSQSGSWCK--TYGEHKHLESHQYVELOR 100
 DB 69 KPNKALRPYTSQSGSWCK--TYGEHKHLESHQYVELOR 109

Search completed: May 6, 2004, 09:04:23
 Job time : 54.614 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:00:50 ; Search time 13.6362 Seconds
(without alignments)
2532.427 Million cell updates/sec

Title: US-09-930-440B-6

Perfect score: 1887

Sequence: 1 MPELEELCPGRWVGQHPGCF.....EEDDTIMEELVDNHGKIKS 359

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1804	95.6	359	2	JC7321
2	619	32.8	337	2	H64432
3	532.5	28.2	341	2	T44651
4	510.5	27.1	346	2	I69836
5	487	25.8	334	2	D81276
6	419	22.2	350	2	F97169
7	382.5	20.3	349	2	S60760
8	370	19.6	373	2	S39722
9	358	19.0	356	2	D87604
10	356.5	18.9	343	2	B81275
11	354	18.8	343	2	H81318
12	303	16.1	340	2	B64542
13	287	15.2	340	2	B71965
14	226	12.0	378	2	H71307
15	122.5	6.5	331	2	B90173
16	119	6.3	332	2	A11899
17	110	5.8	134	2	S53514
18	110	5.8	125	2	B75161
19	108.5	5.7	351	2	S74729
20	108.5	5.7	633	2	D64222
21	108	5.7	63	2	S04974
22	108	5.7	338	2	E72388
23	107.5	5.7	337	2	A97010
24	107	5.7	276	2	E72643
25	106.5	5.6	478	2	B70342
26	102.5	5.4	1140	2	S73786
27	102	5.4	63	2	S04973
28	101	5.4	1107	1	S52517
29	101	5.4	1175	2	T20346

30	100.5	5.3	481	2	H69593	6-phospho-beta-glu
31	100	5.3	1109	2	S53601	myosin-1C [similar
32	98.5	5.2	371	2	A39198	SASP degradation s
33	96.5	5.1	454	2	A43501	sucrose-6-phosphat
34	96.5	5.1	1356	2	T16754	hypothetical prote
35	96	5.1	880	2	F71652	pyruvate, phosphate
36	95.5	5.1	513	2	A05029	reticulocyte-bind
37	95.5	5.1	2829	2	A42771	phosphoenolpyruvat
38	94	5.0	64	1	FDR1RE	antifreeze protein
39	94	5.0	312	2	D90459	conserved hypothet
40	94	5.0	435	2	G96431	probable oxidoredu
41	93.5	5.0	329	2	H83099	antifreeze protein
42	93	4.9	64	2	S53512	cyclohexanone mono
43	93	4.9	542	2	A28550	topoisomerase IV s
44	92.5	4.9	642	2	D90558	hypothetical prote
45	92.5	4.9	1258	2	T29041	

ALIGNMENTS

RESULT 1

JC7321

N-acetylneuraminic acid 9-phosphate synthase (EC 4.1.3.-) - mouse

C/Species: Mus musculus (house mouse)

C/Update: 08-Sep-2000 #sequence revision 08-Sep-2000 #text_change 06-Oct-2000

R/Nakata, D.; Close, B.E.; Colley, K.J.; Matsuda, T.; Kitajima, K.

Biochem. Biophys. Res. Commun. 273, 642-649, 2000

A/Title: Molecular cloning and expression of the mouse N-acetylneuraminic acid 9-phosph

A/Reference number: JC7321

A/Accession: JC7321

A/Molecule type: mRNA

A/Residues: 1-359 <NA>

A/Cross-references: DDBJ:AB041263

C/Comment: This enzyme is a cytosolic enzyme and as a housekeeping enzyme as well, whic

C/Genetics: ceryl-D-mannosamine 6-phosphate to synthesize N-acetylneuraminic acid 9-phosphate.

A/Name: Sialic acid

A/Suprafamily: Streptococcus agalactiae capsular polysaccharide biosynthesis protein cpi

C/Keywords: carbon-carbon lyase; oxo-acid-lyase

Query Match 95.6%; Score 1804; DB 2; Length 359;

Best Local Similarity 94.2%; Pred. No. 1e-133;

Matches 338; Conservative 11; Mismatches 10; Indels 0; Gaps 0;

QY	1	MPELEELCPGRWVGQHPGCFIIAETIGONHOGDLVAKMIRAKGADCAFOKSELEF	60
DB	1	MPELEELCPGRWVGQHPGCFIIAETIGONHOGDLVAKMIRAKGADCAFOKSELEF	60
QY	61	KENKRALSRPYTSKSGMGKTYGEHKKLEFSDOYRELQVLAEEVGIFFTAGGEMAVE	120
DB	61	KENKRALSRPYTSKSGMGKTYGEHKKLEFSDOYRELQVLAEEVGIFFTAGGEMAVE	120
QY	121	FLHELVNPPFKVGSQDTNNPYLEKTAKKGRMVISGQSMQNTKQOVQIYKPLNPNFC	180
DB	121	FLHELVNPPFKVGSQDTNNPYLEKTAKKGRMVISGQSMQNTKQOVQIYKPLNPNFC	180
QY	181	FIQCTSAVLOPEDMNIVISEYOKLFPDIPIGSGHETGINSVAVALGAKVLERHT	240
DB	181	FIQCTSAVLOPEDMNIVISEYOKLFPDIPIGSGHETGINSVAVALGAKVLERHT	240
QY	241	LDKTKGSDHSASLPPGELAEIVRSVRLVERALGSPYKQLLPCMAENCKLAKSVAYKX	300
DB	241	LDKTKGSDHSASLPPGELAEIVRSVRLVERALGSPYKQLLPCMAENCKLAKSVAYKX	300
QY	301	IPAGTTLTDLTLTVAGVEKGYPPEDIFNLAKKVLVTEEDIVMESEVESHKKIKX	359
DB	301	IPAGTTLTDLTLTVAGVEKGYPPEDIFNLAKKVLVTEEDIVMESEVESHKKIKX	359

RESULT 2

N-acetylneuraminic acid synthetase (EC 4.1.3.-) Cj1327 [imported] - Campylobacter jejuni
 C:Species: Campylobacter jejuni
 C>Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 03-Jun-2002
 C:Accession: D81276
 R:Perkhill, J.; Wren, B.W.; Mungall, K.; Kelsey, J.M.; Churcher, C.; Basham, D.; Chiller, G.; Ouali, M.; Rajandream, M.A.; Rutherford, K.M.; VanVlier, A.; Whitehead, S.; Barral, N.M.; Quail, M.; 665-668, 2000
 A>Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp
 A:Reference number: A81250, MID:20150312; PMID:10688204
 A:Accession: D81276
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1334 <PAR>
 A:Cross-references: GB:AL139078; GB:AL111166; NID:96968723; PIDN:CA873754.1; PID:9696876
 A:Experimental source: serotype O2, strain NCTC 11168
 C:Genetics:
 A:Gene: neuB2; Cj1327
 C:Superfamily: Streptococcus agalactiae capsular polysaccharide biosynthesis protein cps
 C:Keywords: carbon-carbon lyase; oxo-acid-lyase

Query Match 25.8%; Score 487; DB 2; Length 334;
 Best Local Similarity 35.1%; Pred. No. 1.6e-30;
 Matches 118; Conservative 64; Mismatches 138; Indels 16; Gaps 5;

QY 21 TIAEIGNHGGDDVAKRMIRMAKCGADCAKFCSELEFRNKALEPPTSKHSG-K 79
 DB 6 IIEAGVNHGGDDVAKRMIRMAKCGADCAKFCSELEFRNKALEPPTSKHSG-K 79
 QY 80 TYGHHKHLFSDQYELQRYAEVGIFFTAGSMDMAVEFLHNLVPEFVKGSGDTNN 139
 DB 66 SQQWQKLELDKAKHKLILHAKKCNIAELSTPFDLESVDLNEGLKFKIPSGITV 125
 QY 140 FPIYLEKAKKGRPMVSSGQSGMDTKQVQIV---KPLNPNCFQCTSAAYLQEDN 136
 DB 126 LPYIKTKAKNKKILITGTANLGEIEBALNVCCKGAKRQNTLTLCITETYPAPFENVA 185
 QY 197 LRVISEYQKLPDIPFGYSGHEGTALISVAVALGATVLERHITLDTKWSGDSHSLSP 256
 DB 186 LKQMOSLKMAF-KLDVGYSHHTGHIHSLAVALGACVIEKHPTLDKMGSPHKSLEP 244
 QY 257 GELAEVRSVRLVERALGSPKQLPCEMACNEKLGSVYAKIPEGTILTMDMITVKY 316
 DB 245 QELMQLCTQIROLKAMGDGKAKSKSEOKNINIVRSILVAKDKIKGEIFSGNLTTK- 303
 QY 317 GEPKAYPEPDI-----NLWGKVLVTEEDDTIME 347
 DB 304 -----RANGISMRVPEFLKATATKTKEDDLIRE 334

RESULT 6
 F97169
 A:Title: Clostridium acetobutylicum
 C:Species: Clostridium acetobutylicum
 C>Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 30-Sep-2001
 C:Accession: F97169
 R:Noelling, J.; Breton, G.; Omelchenko, M.V.; Markova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daily, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
 J. Bacteriol. 183, 4823-4838, 2001
 A>Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium C
 A:Reference number: A96900; PMID:11359325; PMID:21359325
 A:Accession: F97169
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-350 <KUR>
 A:Cross-references: GB:AE001437; PIDN:AA80145.1; PID:G15025183; GSPDB:GN00168
 A:Experimental source: Clostridium acetobutylicum ATCC824
 C:Genetics:
 A:Gene: CAC2187
 C:Superfamily: Streptococcus agalactiae capsular polysaccharide biosynthesis protein cps

Query Match 22.2%; Score 419; DB 2; Length 350;
 Best Local Similarity 32.3%; Pred. No. 3.6e-25;
 Matches 112; Conservative 73; Mismatches 132; Indels 30; Gaps 10;

QY 10 GEMVGGHPCFIITAEIQNHQEDLVAKRMIRMAKCGADCAKFC-----KSELEF 60
 DB 9 GKRIIGGQRTFLIAEMSANHNDPPRAVEIIIAKRSAGDAKLTQYTPDITTPSDNAY 68
 QY 61 KENRKALERPYSKSKSWGTYGKHLEFSDQYR-ELQRYAEVGIFFTAGSMDMAV 119
 DB 69 POLKQGTITWDGTTLH---KLVE-----AYTWQOPKLEIAEBGLCLFSSPNTSV 120
 QY 120 EELAEMLVPPFFVSGSDTNNPYLEKTKAKGRPMVSSGQSGMDTKQVQIV---KPL 175
 DB 121 DELEKKEVAPYKVASFELTDIPFIEYIAKGRPMVSSGQSGMDTKQVQIV---KPL 177
 QY 176 NENFCLQCTSAAYLPQPEVNLKRVISEVQKLPDIPFGYSGHEGTALISVAVALGAKYL 235
 DB 178 NENVILKCSSYSPPELEINLKTIPNNKEAF-NCVGLSDITMGYSATPAVALDAIYI 236
 QY 236 EKHITLDTKWSGDSHSLSPGELAEVRSVRLVERALGSPKQLPCEMACNEKLGKGV 295
 DB 237 EKHITLDTKWSGDSHSLSPGELAEVRSVRLVERALGSPKQLPCEMACNEKLGKGV 295
 QY 296 VAKVKIPEGTILTMDMITVKVGEPR-AIPPEDIPLVCKVLYVYVE 341
 DB 296 FVVKIKKGEFTTKR--NVKSIKPAFGILTKYIEVIGKRAVDYVK 340

RESULT 7
 S60760
 A:Title: Molecular analysis of the biosynthesis pathway of the alpha-2,8 polysialic acid
 A:Reference number: S60758; PMID:9531727; PMID:783052
 A:Accession: S60760
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-349 <EDM>
 A:Cross-references: EMBL:M95053; NID:9520732; PIDN:AA20477.1; PID:9520735
 A>Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1994
 R:Teitelin, H.; Saunders, W.U.; Heideberg, J.; Jeffries, A.C.; Nelson, K.B.; Bisen, J.;
 Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;
 H.; Qin, H.; Vanatavan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizze, M.
 Science 287, 1809-1815, 2000
 A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappelli, R.; V
 A>Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
 A:Reference number: A81000; PMID:10715755; PMID:10710307
 A:Accession: B81241
 A:Molecule type: DNA
 A:Residues: 1-349 <DET>
 A:Cross-references: GB:AE002366; GB:AE002098; NID:97225284; PIDN:AA40535.1; PID:9722528
 A:Experimental source: serogroup B, strain MC58
 C:Genetics:
 A:Gene: AMB0068
 C:Superfamily: Streptococcus agalactiae capsular polysaccharide biosynthesis protein cps

Query Match 20.3%; Score 382.5; DB 2; Length 349;
 Best Local Similarity 30.5%; Pred. No. 2.6e-22;
 Matches 93; Conservative 57; Mismatches 152; Indels 3; Gaps 1;

QY 11 RWVGGHPCFIITAEIQNHQEDLVAKRMIRMAKCGADCAKFCSELEFRNKALEP 70
 DB 12 RSVGNHEPLIICEIGINIEGSLKTFEVDVAVYAGAVVHQHYIVDEHSDKQ-- 69
 QY 71 YTSKSWKTYGKHLEFSDQYR-ELQRYAEVGIFFTAGSMDMAVFLHNLVPEF 130
 DB 70 -VLPNADVSIYERICALNEDEIKLRYESKMPISIPFSRAALRIQMDIDIPAY 128
 QY 131 KVGSGDTNNPYLEKTKAKGRPMVSSGQSGMDTKQVQIVKPLNPNCFQCTSAAYL 190
 DB 129 KVGSGCENYPIPKLVASFGEKILITSTGKNSIESIKSVETIRAGVPYALHCTNITFT 188

QY 191 QPEDVNLRVISRYOKLFPDIPIGYSGHETGIAISVAANLGAUKLERHTITLDMKWSGDH 250
 DB 189 PYEDVRLGANDLSEFPDPAITIGLSOHTIDNTACIAQVALGSLIERHTIDMRPRPDI 248
 QY 251 SASLPEGLAEIVRSVRLVERALGSEPTKOLLPCENACNEKLGKSVAVAKVIEGTITLMD 310
 DB 249 VCSMPDPTKEKLGQAGHALKARGKQDITTAGKRTKOPAPASVADKIDKGLSD 308
 QY 311 MLYTK 315
 DB 309 NLWVK 313
 RESULT 8
 539722
 spore coat polysaccharide synthesis protein spse - Bacillus subtilis
 N:Alternate names: protein ipa-66d
 C:Species: Bacillus subtilis
 C:Date: 07-Oct-1994 #sequence_revision 26-May-1995 #text_change 20-Jun-2000
 C:Accession: S39722; C69717
 R:Glasner, P.; Kunst, F.; Arnold, M.; Coudart, M.F.; Gonzales, W.; Hullo, M.F.; Ionescu, A.; Rapoport, G.; Danchin, A.
 Mol. Microbiol. 10, 371-384, 1993
 A>Title: Bacillus subtilis genome project: cloning and sequencing of the 97 kb region fr
 A:Reference number: S39655; PMID:95020537; PMID:7934828
 A:Accession: S39722
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-373 <DNA>
 A:Cross-references: EMBL:X73124; NID:g413923; PIND:CA51623.1; PID:g413951
 A>Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1993
 R:Kunst, F.; Ogatawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
 C.; Bron, S.; Brulliet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chd
 A.; Ehrlich, S.D.; Emmerison, P.T.; Entian, K.D.; Errington, J.; Faltret, C.; Ferrati, E.
 Nature 390, 249-256, 1997
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizi, A.; Gall
 tech, J.; Harwood, C.R.; Henauf, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
 Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
 A.; Authors: Labber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mauei
 Y., M.; Ogawa, K.; Ogimura, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetel
 Rieger, M.; Rivolta, C.; Rocha, B.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon,
 A.; Authors: Schleich, S.; Schroeder, R.; Scoffone, F.; Sekiguchi, T.; Sekowska, A.; Seron
 T.; Willems, P.; Wipert, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
 A.; Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
 A>Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
 A:Reference number: A69580; PMID:98044033; PMID:9384377
 A:Accession: C69717
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-373 <DNA>
 A:Cross-references: GB:299123; GB:AL009126; NID:g2636240; PIND:CAB15813.1; PID:g2636322
 A:Experimental source: strain 168
 C:Genetics:
 A:Gene: spse
 C:Superfamily: Streptococcus agalactiae capsular polysaccharide biosynthesis protein cps

Query Match 19.6%; Score 370; DB 2; Length 373;
 Best Local Similarity 30.1%; Pred. No. 2,7e-21;
 Matches 107; Conservative 55; Mismatches 167; Indels 26; Gaps 9;
 QY 13 VGGQHPCFIATIGONHGDIVAKMIRPAKCGADCAFPKSELEFKERKALRPPYT 72
 DB 11 VGGDAVFITIAEGINHDKLQAFALIDAAAGADAVAFQY----FQADRYVQKDPGL 66
 QY 73 SKRSWGKTYGZHK--RHLEFSHQYRELQRYAEVGIFFPASGDMAYEFLEHNVPPF 130
 DB 67 YKTAAGCDVSIFFSLVQSWMPAMWILPLLDYCREKQVITFLSTVCDGASDLQSTPSAPF 126
 QY 131 KVSQSDTNPPVLEKTKKGRPMVVISGQSMQMDTKQVQIVPL--NPFFCFLOCTSAAP 189
 DB 127 KIASYEINHPLKLYVRLRRLRPFSTAGAEISDVHEAWRTTAAEENQAIWECAKXP 186
 RESULT 9
 87604
 neub protein, probable [imported] - Caulobacter crescentus
 C:Species: Caulobacter crescentus
 C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
 C:Accession: D87604
 R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J
 B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolc
 n, J.; Esmolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.H
 Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
 A>Title: Complete Genome Sequence of Caulobacter crescentus.
 A:Reference number: A87249; PMID:21173698; PMID:11259647
 A:Accession: D87604
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-356 <STO>
 A:Cross-references: GB:ABD05673; NID:q1342484; PIND:AAK24832.1; GSPDB:GN00148
 C:Genetics:
 A:Gene: CC2868
 C:Superfamily: Streptococcus agalactiae capsular polysaccharide biosynthesis protein cf
 Query Match 19.0%; Score 358; DB 2; Length 356;
 Best Local Similarity 30.1%; Pred. No. 2.3e-20;
 Matches 102; Conservative 55; Mismatches 170; Indels 12; Gaps 9;
 QY 1 MFLBELCPGRKVGQHPCFIATIGONHGDIVAKMIRPAKCGADCAFPKSELEF 60
 DB 8 IFFSIEIA--GRKIGADHPVYICELSGNLSLEKLAADVADTGDAIKITQYTD-- 65
 QY 61 KENRALRPRYTKS--W--GKTYGHRHLEFSHQYRELQRYAEVGIFFPASGDMEM 117
 DB 66 -TITLDVRPEKIHGIMQRTLYELYEAKHTPEEMAAIFERRARGVITFFSPDET 124
 QY 118 AVEPLHEINVPPEKVGSDTNPPVLEKTKKGRPMVVISGQSMQMDTKQVQIVPLN- 176
 DB 125 AVDLIDSGAPFAFKTASFEAVDLPLIKYAAKQXLLISTGMAMTMOALDTLSAGA 184
 QY 177 EMFGLGCTSAVPLPEPVNLRVISRYOKLFPDIPIGYSGHETGIAISVAANLGAUKLE 236
 DB 185 FGVLLHACVSTFAPAPAAVTVTDMAARF--GCITGSDHTRCGASVAANSLGACAV 243
 QY 237 RHITLDMKWSGDHSALEPGLAEIVRSVRLVERALGSEPTKOLLPCENACNEKLGKSV 296
 DB 244 KHTPLRADGGPDFAFSLPEAFKALVDTDKVAMALRAHYDVGSF--ATSLFRRLSY 302
 QY 297 AKVTLPEGTITL--MDMLTVKQGEKAPVPEPFIWVGK 334
 DB 303 VTADVKAEPILTRAVRSVRG--NGLPADLDKVLAK 339
 RESULT 10
 B81275
 N-acetylneuraminic acid synthetase (EC 4.1.3.-) Cj1317 [imported] - Campylobacter jejuni
 C:Species: Campylobacter jejuni
 C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 03-Jun-2002
 C:Accession: B81275
 R:Parthill, J.; Wren, B.W.; Mungall, K.; Kestley, J.M.; Churcher, C.; Basham, D.; Chilli
 C.W.; Quill, M.; Rajadream, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Barre
 Nature 403, 665-666, 2000

A>Title: The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hyl
 A:Reference number: A61250; M01D:20150912; PMID:10688204
 A:Accession: B81275
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-343 <PAR>
 A:Cross-references: GB:AL139078; GB:AL111168; NID:96968723; PIDN:GAB73744.1; PID:9696872
 A:Experimental source: serotype O2, strain NCTC 11168
 A:Genetics: newB3; Cj1317
 C:Superfamily: Streptococcus agalactiae capsular polysaccharide biosynthesis protein cps
 C:Keywords: carbon-carbon lyase; oxo-acid-lyase

Query Match 18.9%; Score 356.5; DB 2; Length 343;
 Best Local Similarity 29.9%; Pred. No. 2.8e-20;
 Matches 104; Conservative 65; Mismatches 138; Indels 41; Gaps 10;

20 FIIAELGQNHQDLDVAKRMIRAKGCGACAFQ-----KSELERK---FKR 64
 13 FIIAELSANHAGSLKALSKAKKAGADAIKIOTYTPDSLTLNDSKEDFTIKGLMDK 72
 65 KALERPYS---HSMGKTGGEKRLHPSHDQYRELQRYAEVGIFFTASGMDKAVAF 121
 73 RKLYEYSKATYEM-----HSQIFET---AONEGLICSSPPAKEDVDF 115
 122 LHEINVPFFVYSGDNNPPEYLSKTAKEPWISSGMSQMDTMKQVQIVK-PLNPNC 180
 116 LKAFDILAYKIAFENANDENFRLIKKPKTIVSTGITEBELFICELFEEKNPDLY 175
 181 FLOCTAVPLQEPDNLRVISEQKIFPDIPIGSGEHTGIAISVAVALGAKVLEHRT 240
 176 ELKCTGYTPAIDMMLKIGVSKERF-NVEGLSHSREFLAPVAVALGARVIRKHEM 234
 241 LDKTWKSGDHSALREBELAIRVSVRLVERALGSPTKOLLPCENAKNEKGVAAKYK 300
 235 LDKSISEDSKFLDPDEFKAMDAVROAESALGD-GKLDLDEKVLKNRFAASLYASXD 293
 301 IPEGTILTMMLTVKGEPR-AAPPEDIFLVNKKLVVYEHDDTIME 347
 294 IKXGEMFSES--NKSVPSPFGILHPKYQELLGNKASKDKIDFDALQ 339

RESULT 11
 H81318
 N-acetylneuraminic acid synthetase (EC 4.1.3.-) Cj1141 (imported) - *Campylobacter jejuni*
 C:Species: *Campylobacter jejuni*
 C>Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 03-Jun-2002
 C:Accession: H81318
 R:Parkhill, J.; Wren, B.W.; Mungall, K.; Kelsey, J.M.; Churcher, C.; Basham, D.; Chilling
 C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Barrett
 Nature 403, 665-668, 2000
 A>Title: The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hyl
 A:Reference number: A61250; M01D:20150912; PMID:10688204
 A:Accession: H81318
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-343 <PAR>
 A:Cross-references: GB:AL139077; GB:AL111168; NID:96968444; PIDN:GAB73396.1; PID:9696857
 A:Experimental source: serotype O2, strain NCTC 11168
 C:Genetics:
 A:Gene: newB1; Cj1141
 C:Superfamily: Streptococcus agalactiae capsular polysaccharide biosynthesis protein cps
 C:Keywords: carbon-carbon lyase; oxo-acid-lyase

Query Match 18.8%; Score 354; DB 2; Length 343;
 Best Local Similarity 30.5%; Pred. No. 4.4e-20;
 Matches 96; Conservative 57; Mismatches 156; Indels 6; Gaps 3;

21 IIAELGQNHQDLDVAKRMIRAKGCGACAFQSKSELEFRKALE-REYTSKHSWGK 79
 17 IIEELGINHNSLEIAKLMDAAKRAKAKIKIOTHTIVEDMSQEAQKIVPGNANIS--- 73
 80 TYGKHKHLFSPHQYRELQRYAEVGIFFTASGMDKAVAFELHEINVPFFVYSGDNN 139

DB 74 -IYEIMEQCALNYKDELALKEVEKQGLVYLSTPSPRAAANRLBDMGSAVYKISGGECCNN 132
 QY FRYAEKTKKGRPMWISSGMSQMDTMKQVQIVKPLNPNFCLOCTSAVPLQEPDNLRV 199
 DB 133 YPLHILHIOFKPKPMIISTGMSISEIKFTVILIRYELPFLVILHTNLNLYPTPSHVRLOA 192
 QY ISEYOKLFPDIPIGSGEHTGIAISVAVALGAKVLEHRTDKTWKSGDHSALREPEL 259
 DB 193 MLELYER-NCLYGSDDHTNNALCIGALALGASVLEHRTDTMDRKQPDIVCSMDSESTL 251
 QY AELVSVLVERALGSPTKOLLPCENAKNEKGVAAKYVPIEGTILTMMLTVKGEPR 319
 DB 252 KDLINQTEMLVLRDNNKNPLKEQVTTDPAPASVSIKDIKSGEILSMONIVYKRSK 311
 QY 320 KAYPPEDIFNLVGRK 334
 DB 312 GGISAKDPEALIGKR 326

RESULT 12
 B64542
 Spore coat polysaccharide biosynthesis protein E - *Helicobacter pylori* (strain 26995)
 C:Species: *Helicobacter pylori*
 C>Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 20-Jun-2000
 C:Accession: B64542
 R:Tom, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Piechmann, R.D.
 Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khaliq, H.G.; Glodex, A.; McKen
 son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Wathey, L
 Nature 388, 539-547, 1997
 A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C
 A>Title: The complete genome sequence of the gastric pathogen *Helicobacter pylori*.
 A:Reference number: A64520; M01D:97394467; PMID:9252185
 A:Accession: B64542
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-340 <Tom>
 A:Cross-references: GB:A600538; GB:A600511; NID:92313263; PIDN:AAD07248.1; PID:9231326
 C:Superfamily: Streptococcus agalactiae capsular polysaccharide biosynthesis protein cps

Query Match 16.1%; Score 303; DB 2; Length 340;
 Best Local Similarity 27.7%; Pred. No. 4.3e-16;
 Matches 91; Conservative 66; Mismatches 149; Indels 22; Gaps 10;

18 PFTIAELGQNHQDLDVAKRMIRAKGCGACAFQK-----SELEFRKALERPYT 72
 4 PFTIAELSANHAGSLKALSKAKKAGADAIKIOTYTPDSLTLNDSKEDFTIKGLMDK 63
 73 SKHSWGKTYGKRLHPSHDQYRELQRYAEVGIFFTASGMDKAVAFELHEINVPFFRV 132
 DB 64 DENNYELQKASTLEW-HAELELAKTL-DLGIF--SSPSSQALTELESLANCPMYE 119
 QY 133 GSGDNNPPEYLSKTAKEPWISSGMSQMDTMKQVQIVKPLNPNF---CLOCTSAVPL 190
 DB 120 ASPEYVDLLEKARQEPITLSSGIALHTHTLOAISICRYV-NFDITLAKCSAYIS 178
 QY 191 QPEDVNLRVISEYOKLFPDIPIGSGEHTGIAISVAVALGAKVLEHRTDKTWKSGDH 250
 DB 179 KIEDANLISMVGLGELF-GVRFGLSDHTIGSLPIATLTIGASMEIKHFLINKSLQTPPS 237
 QY 251 SASLEPGLAEIVSVRLVERALGSPTKOLLPCENAKNEKGVAAKYKIDEGTILF-- 308
 DB 238 ASKMPFNGKSVVAKIKOSVIALGSEEPINPTLEKRRFPARSLFVITKDIQGEALTEN 297
 QY 309 -MDMLTVKVG-EPKAYPPEDIFNLVGRK 334
 DB 298 NIKALRPNIGLHPKRYK-----ELLQOK 320

RESULT 13
 B73965
 Sialic acid synthase - *Helicobacter pylori* (strain J99)
 C:Species: *Helicobacter pylori*

A:Variety: strain J99
 C:Date: 12-Feb-1999 #sequence revision 12-Feb-1999 #text_change 20-Jun-2000
 C:Accession: B71965
 R:Alm, R.A.; Ling, L.S.; L., Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;
 Ives, C.; Gibson, R.; Merberg, D.; Miller, S.D.; Jiang, Q.; Taylor, D.E.; Voyts, G.F.;
 Nature 397, 176-180, 1999
 A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric path
 A:Reference number: A71800; PMID:99120557; PMID:9923682
 A:Accession: B71965
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-340 <ARN>
 A:Cross-references: GB:AE001455; GB:AE001459; NID:G4154678; PIDN:AAD05747.1; PID:G415468
 A:Experimental source: strain J99
 C:Genetics:
 A:Gene: neuB
 C:Superfamily: Streptococcus agalactiae capsular polysaccharide biosynthesis protein cps

Query Match 15.2%; Score 287; DB 2; Length 340;
 Best local similarity 27.8%; Pred. No. 7, 8e-15;
 Matches 92; Conservative 68; Mismatches 137; Indels 34; Gaps 14;

21 TTAETGQNHQGDIDVAKRMIRAKKCGACAKPQK-----SELEFRPKALERPYTSN 75
 7 IVAELSANRQDNLAKESLHAIKESGADFVXLQYTPSCMTLDSK-----EDPPIIQG 60
 76 S-WGK--TYGHRK---HLEFSDQYRELQRYAEVIGIFFTASGNDMAVEFLHEINVPF 129
 61 TMDQENLYGLYOKASTPLEW-HAELFELAKL-DLGI--SSPSSKALLESLSDCPM 116
 130 FVYSGDITNNPYLETAKKRPMTVSSGMSQMDTKQYQVPLNPF--CFLOCTSA 187
 117 YKIASFEIYDLDLIEKARTQKPIILSSGIAHAEIDASISCRGN-NFDITLLKCYSA 175
 188 YLQPEDVNLRYISEYQKFPDIPICYSHERGIAISVAALGAKVLEHRTLDKTMWG 247
 176 YSKEDDAHLKMKVKGERT-GVKFLSHHTIGSLPILATLIGASMKHEILNLSOT 234
 248 SPHSASLEPGEALVRSVLEFALGSPTKQLLPCEMACNEKLGKSVAKYKIPGTL 307
 235 PSASFMDTFNGFSKMGAIKQSVLAGEEKPINFTLEERRFARSLFVIDIQGAL 294
 308 TMD--MLTVKVG-EKAYRPEDIFNLGKK 334
 295 TSDNIRALRPNDLGFPRYK-----EILDQK 320

Db

RESULT 14
 H71307
 Probable spore coat polysaccharide biosynthesis protein (spse) - syphilis spirochete
 C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)
 C:Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 19-Apr-2002
 C:Accession: H71307
 R:Fraser, C.M.; Norris, S.J.; Weinstein, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwin
 rson, J.; Khilak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; McD
 Science 281, 375-388, 1998
 A:Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.
 A:Reference number: A71250; PMID:9832770; PMID:9665876
 A:Accession: H71307
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-378 <COL>
 A:Cross-references: GB:AB001232; GB:AB000520; NID:G3322856; PIDN:AAC6539.1; PID:G332285
 A:Experimental source: strain Nichols
 C:Genetics:
 A:Gene: TPO562
 C:Superfamily: Streptococcus agalactiae capsular polysaccharide biosynthesis protein cps

Query Match 12.0%; Score 226; DB 2; Length 378;
 Best local similarity 25.1%; Pred. No. 5, 4e-10;
 Matches 85; Conservative 47; Mismatches 166; Indels 40; Gaps 9;

8 CQSWGGQHPFCFIETGQNHQGDIDVAKRMIRAKKCGACAKPQKSELEFRPKAL 67
 4 CGGCRPADIDILTAIGASHGSPRABALIDAAADAAAVAKQ-----LIVAEI 57
 68 ERYPTK--HSGWKYGEHRKHLFESHQYRELQRYAEVIGIFFTASGNDMAVEFLHE 124
 58 LHPITGAVRLPSGAVSVSYQREELVPLSFYAOQFNARSRGMLVGISPPGSRATATA 117
 125 LNVPPFYVSGDITNNPYLEKT-AKGRPMVSSGM-----QSMDTKQYQVPLNPF 179
 118 LKDFPLVASEL-NYTLISTIAAALPLISSGVCGLKEIBGALAQCQYTKQ-GSSH 175
 180 CFLOCTSAVPLQPEDVNLRYISEYQKLPDIPICYSHERG-IAISVAALGAKVLEH 238
 176 ALHICITAYPAPERTRYMALPALATIF-NINVSQHSVDPLVPLARHGAQYVEKH 234
 239 ITLDTKMGSDHSASLEPGEALVRSV-----RLVERALGSP 276
 235 ICLSRIDAGLDDSTALDPADERTMTALNSCARSPQIISFLHERGAYPVAVAVIGSG 294
 277 TKQLPCEMACNEKLGKSVAKYKIPGTLITMDMLTV 314
 295 EKVLAPSERAHYOKSNRSIHYLHAYPRGTVALQKENLVI 332

Db

RESULT 15
 B80173
 Hypothetical protein aroG [imported] - Sulfolobus solfataricus
 C:Species: Sulfolobus solfataricus
 C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 18-Nov-2002
 C:Accession: B80173
 R:She, Q.; Singh, R.K.; Contaloni, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Char
 Jorg, I.; Jeffries, A.C.; Kozera, C.O.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder,
 arett, R.A.; Ragan, M.A.; Sersen, C.W.; Van der Oost, J.
 submitted to GenBank, April 2001
 A:Description: Sulfolobus solfataricus complete genome.
 A:Reference number: A99139
 A:Accession: B80173
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-331 <KUN>
 A:Cross-references: GB:AE006641; NID:G13813447; PIDN:AAK40641.1; GSPDB:GN00155
 C:Genetics:
 A:Gene: aroG
 C:Superfamily: phospho-2-dehydro-3-deoxyoctonate aldolase

Query Match 6.5%; Score 122.5; DB 2; Length 331;
 Best local similarity 23.9%; Pred. No. 0.059;
 Matches 66; Conservative 46; Mismatches 103; Indels 61; Gaps 15;

18 PCFIETGQNHQGDIDVAKRMIRAKKCGACAKPQKSELEFRPKALERPYTSKSW 77
 95 PCAV-----HNEBQYLTAKAV-----KRGASL-----RGAYKPRTPSPSF 133
 78 KTYGEHRKHLFESHQYRELQRYAEVIGIFFTASGNDMAVEFLHEINVPFKGSDT 137
 134 -GGLGE-----EGVKILRRVQDEVGALPIVEIMDTRDSNIFSQY-VMIOIGASNA 182
 138 NNFPYLETAKKRPMTVSSGMSQMDTKQYQVPLNPF-----NPNFC-----FLOCTSA 186
 183 QNFSLIKVGLGKFLKRGNG-NVBEWLOAAEYILLEGNNNTVCEGIRTFEEST 240
 187 AYVLPQPEDVNLRYISEYQKLPDIPY-GYSGHERG-----IAISVAALGAKVLEH 238
 241 RFTL--DIGGVAA--KLMTHTLPICADPSHPGKREIVSHLAAVAAAGADMILLVH 294
 239 ITLDTKMGSDHSASLEPGEALVRSVLEFALG 274
 295 PHEKAL--SDSQQLTPESFEVLMNRITLAKALG 328

Db

Search completed: May 6, 2004, 09:09:45
 Job time: 16.8862 secs

Thu May 13 11:53:10 2004

us-09-930-440b-6.rpt

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OW protein - protein search, using SW model

Run on: May 6, 2004, 08:56:04 / Search time 9.44047 Seconds
(without alignments)
1980.112 Million cell updates/sec

Title: US-09-930-440b-6

Perfect score: 1887
Sequence: 1 MFLELELCPRGHWG3QHPCF.....EEDDTIMEELVDNKGKIKS 359

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 141681 segs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seg length: 0
Maximum DB seg length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB ID	Description
1	1883	99.8	359	1	SIMS_HUMAN
2	619	32.8	337	1	YAES_METUA
3	370	19.6	373	1	SPSE_BACST
4	113.5	6.0	1248	1	SMIB_MOUSE
5	110	5.8	134	1	ANP3_RHIDE
6	108.5	5.7	633	1	PAR2_WYCGE
7	108	5.7	63	1	ANP2_AUSBR
8	108	5.7	338	1	AROF_THEMA
9	106.5	5.6	478	1	GATB_AQUAS
10	102.5	5.4	1140	1	YD76_MYCPN
11	102	5.4	63	1	ANP1_AUSBR
12	101	5.4	1107	1	MYLE_RAT
13	100.5	5.3	469	1	BGL1_BACST
14	100	5.3	1109	1	WYLE_HOMAN
15	98.5	5.2	371	1	GPR_BACME
16	97.5	5.2	642	1	DNAR_XANCP
17	97	5.1	378	1	ARGE_VIBIO
18	96.5	5.1	641	1	DNAR_XANAC
19	96.5	5.1	1437	1	DP03_LACPL
20	96	5.1	880	1	POK8_RICPR
21	95.5	5.1	513	1	CHAB_MARPO
22	95.5	5.1	2869	1	RBP1_PLAAB
23	94	5.0	512	1	ANP2_RHIDE
24	93.5	5.0	64	1	GPW1_CLOPE
25	93	4.9	64	1	ANP1_RHIDE
26	93	4.9	542	1	CYVO_ACIAP
27	92.5	4.9	1333	1	YP74_CABEU
28	92	4.9	346	1	BPH1_BURCE
29	92	4.9	638	1	DNAR_XYLF
30	92	4.9	1656	1	ATG8_YEAST
31	91.5	4.8	458	1	EF12_RHIRA
32	91.5	4.8	458	1	EF12_RHIRA
33	91	4.8	91	1	ANP3_MACAM

34	91	4.8	197	1	CLPP_BUCAP	Q8K990	buchnera ap
35	90	4.8	538	1	CP18_DROME	Q85078	drosophila
36	89.5	4.7	381	1	NCAP_CVCAI	P36298	canine ente
37	89.5	4.7	397	1	LE11_METKA	Q81W28	methanopyru
38	89.5	4.7	593	1	MDLB_BCOLI	P75706	escherichia
39	89.5	4.7	638	1	DNAR_XYLF	Q87B58	xyliella fas
40	89	4.7	838	1	ETPL_XANCP	P45597	x multiphos
41	88.5	4.7	387	1	YNV8_YEAST	P40151	seccharomyc
42	88	4.7	66	1	ANP1_LYCPO	P24028	lycodes pol
43	88	4.7	479	1	SCRB_STIRU	P13522	streptococc
44	88	4.7	508	1	CHLB_SYMP7	Q8J3N0	synechococc
45	88	4.7	855	1	ORP3_MOUSE	Q9db59	mus musculi

ALIGNMENTS

RESULT 1	SIMS_HUMAN	STANDARD	PRT	359 AA.
ID	SIMS_HUMAN	STANDARD	PRT	359 AA.
AC	Q8NR45; Q8NRV9; Q9BWS6; Q9NV04;			
DT	28-FEB-2003 (Rel. 41, Created)			
DT	28-FEB-2003 (Rel. 41, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	acetylneuraminic acid synthase (N-acetylneuraminic acid synthase) (EC 2.5.1.56) (N-acetylneuraminic acid synthase) (EC 2.5.1.57) (N-acetylneuraminic acid synthase)			
DE	SAS.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.			
OX	NCBI_TaxID:9606;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND CHARACTERIZATION.			
RC	TISSUE=Liver;			
RX	MEDLINE=20298869; PubMed=10749855;			
RA	Lawrence S.M., Huddleston K.A., Pitts L.R., Nguyen N., Lee Y.C.,			
RA	Vann W.F., Coleman T.A., Betenbaugh M.J.;			
RT	"Cloning and expression of the human N-acetylneuraminic acid phosphate			
RT	synthase gene with 2-keto-3-deoxy-D-glycero-D-galacto-nononic acid			
RT	biosynthetic ability."			
RL	U. Biol. Chem. 275:17869-17877 (2000).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RA	Isegai T., Oca T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,			
RA	Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.;			
RA	Matsunawa H., Ichii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.;			
RA	Nakamura Y., Nagahara K., Masuko Y., Sasaki N.;			
RT	"NEO human cDNA sequencing project."			
RT	Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.			
RN	[3]			
RP	SEQUENCE FROM N.A., AND VARIANT ASP-68.			
RC	TISSUE=Lung, and Placenta;			
RX	MEDLINE=22388257; PubMed=12477932;			
RA	Straussberg R.L., Feinberg F.S., Wagner L., Sherman C.M., Schuler G.D.;			
RA	Klausner R.D., Collins F.S., Schaefer C.P., Bat N.K.;			
RA	Altschul S.F., Zeeberg B., Buetow K.H., Moore T., Wax S.I., Wang J., Hsieh P.;			
RA	Hopkins R.F., Jordan K., Farmer A.A., Rubin G.M., Hong L.;			
RA	Diatchenko L., Marinina K., Palmer M.F., Casavant T.L., Scheetz T.B.;			
RA	Stapleton M.J., Soares M.B., Bonaldo M.F., Carninci P., Prange C.;			
RA	Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Mullany S.J.;			
RA	Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Gunaratne P.H.;			
RA	Bohak S.A., McQueen P.J., McKernan K.J., Malek J.A., Gunaratne P.H.;			
RA	Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Halys S.W.;			
RA	Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.;			
RA	Foley J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.;			
RA	Whiting R., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.;			
RA	Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.;			
RA	Rodriguez A.C., Grimwood J., Schmutz J., Myers R.N.;			
RA	Butterfield J.S.N., Krzywinski M.I., Skalska U., Smallus D.E.;			
RA	Schneerich A., Schein J.E., Jones S.J.M., Marra M.A.;			
RT	"Generation and initial analysis of more than 15,000 full-length			
RT	human and mouse cDNA sequences."			

Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

- FUNCTION: Produces N-acetylneuraminic acid (Neu5Ac) and 2-keto-3-deoxy-D-glycero-D-galacto-nononic acid (KDN). Can also use N-acetylmannosamine 6-phosphate and mannose 6-phosphate as substrates to generate phosphorylated forms of Neu5Ac and KDN, respectively.

- CATALYTIC ACTIVITY: Phosphoenolpyruvate + N-acetyl-D-mannosamine + H₂O = phosphate + N-acetylneuraminic acid.

- CATALYTIC ACTIVITY: Phosphoenolpyruvate + N-acetyl-D-mannosamine 6-phosphate + H₂O = N-acetylneuraminic acid + 9-phosphate + phosphate.

- TISSUE SPECIFICITY: Ubiquitous.

- SIMILARITY: Contains 1 AFP-like domain.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AF257466; AAF75261.1; -
 EMBL; AK001653; BAA91818.1; -
 EMBL; BC000008; AAH00008.1; -
 EMBL; BC019315; AAH19315.1; -
 MIM: 605202; -
 HSP; P19614; NAME.
 GO: GO:0005737; C:cytoplasm; NAS.
 GO: GO:0008781; F:N-acetylneuraminic acid synthetase acti. ; NAS.
 GO: GO:0009103; F:phosphoenolpyruvate biosynthesis; NAS.
 InterPro: IPR006014; Antifreeze_dom.
 InterPro: IPR006190; Antifreeze_like.
 InterPro: IPR004144; Neut.
 Pfam; PF01354; Antifreeze; 1.
 Pfam; PF0102; Neut; 1.
 ProDom; PD003258; Antifreeze1; 1.
 ProSite; PS50844; AFP_LIKE; 1.
 Transferrase; Polymorphism.
 DOMAIN 294 353 AFP-LIKE.
 VARIANT 68 68 E -> D.
 CONFLICT 232 232 /FTID-VAR 013308.
 CONFLICT 321 321 A -> T (IN REF. 2).
 SEQUENCE 359 AA; 40307 MW; G-E02D47F4F98592F CRC64;

Query March 99.8%; Score 1883; DB 1; Length 359;
 Best local similarity 99.7%; Pred. No. 7,2e-136;
 Matches 358; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MELELECPGAWGQHPCTIABIGNHQGDLDVAKMIMAKKCGADCAKPKSELEF 60
 DB 1 MELELECPGAWGQHPCTIABIGNHQGDLDVAKMIMAKKCGADCAKPKSELEF 60
 QY KFNRALEPPTSKSKWKTGEGHREHLEFSHDYREIQRVAEEVGFPTASGDENAVE 120
 DB KFNRALEPPTSKSKWKTGEGHREHLEFSHDYREIQRVAEEVGFPTASGDENAVE 120
 QY 61 KFNRALEPPTSKSKWKTGEGHREHLEFSHDYREIQRVAEEVGFPTASGDENAVE 120
 DB 61 KFNRALEPPTSKSKWKTGEGHREHLEFSHDYREIQRVAEEVGFPTASGDENAVE 120
 QY 121 FLHEINVEFPKVGSGDITNFPYLKRTAKKGRPMVSSGQSGMDTKQYQIVKPLNPF 180
 DB 121 FLHEINVEFPKVGSGDITNFPYLKRTAKKGRPMVSSGQSGMDTKQYQIVKPLNPF 180
 QY 121 FLHEINVEFPKVGSGDITNFPYLKRTAKKGRPMVSSGQSGMDTKQYQIVKPLNPF 180
 DB 121 FLHEINVEFPKVGSGDITNFPYLKRTAKKGRPMVSSGQSGMDTKQYQIVKPLNPF 180
 QY 181 FLQCSAPLPDPEDNLEVISYQCLPDPPIGSGHETGIAISVAALAKVLERIT 240
 DB 181 FLQCSAPLPDPEDNLEVISYQCLPDPPIGSGHETGIAISVAALAKVLERIT 240
 QY 181 FLQCSAPLPDPEDNLEVISYQCLPDPPIGSGHETGIAISVAALAKVLERIT 240
 DB 181 FLQCSAPLPDPEDNLEVISYQCLPDPPIGSGHETGIAISVAALAKVLERIT 240
 QY 241 LDXTKSGDHSASLEPGEIABELVRSVLVERALGSPTRQLPCENACNEKLGKSVANV 300
 DB 241 LDXTKSGDHSASLEPGEIABELVRSVLVERALGSPTRQLPCENACNEKLGKSVANV 300
 QY 301 IPEGITLMDLTVVYGEKAYRPEDINLVAKKLVTVVEEDDTMEILVNEGKTKS 359
 DB 301 IPEGITLMDLTVVYGEKAYRPEDINLVAKKLVTVVEEDDTMEILVNEGKTKS 359
 QY 301 IPEGITLMDLTVVYGEKAYRPEDINLVAKKLVTVVEEDDTMEILVNEGKTKS 359
 DB 301 IPEGITLMDLTVVYGEKAYRPEDINLVAKKLVTVVEEDDTMEILVNEGKTKS 359

RESULT 2
 YAE5 METVA STANDARD; PRT; 337 AA.
 ID YAE5 METVA
 AC 058465;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Hypothetical protein M1065.
 GN M1065.
 OS Methanococcus jannaschii.
 OC Archaea; Buryarchaeota; Methanococci; Methanococcales;
 CC Methanocaldococcaceae; Methanocaldococcus.
 CX NCBI_TaxID=2190;
 [1]
 SEQUENCE FROM N.A.
 RP STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
 RX MEDLINE=96337939; PubMed=8688087;
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
 RA Ketravage A.R., Dougherty B.A., Tomb J.F., Adams M.D., Reich C.I.,
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Gilek A.,
 RA Scott J.L., Geoghegan N.S.M., Feldman J.F., Fuhrman J.L., Nguyen D.,
 RA Ustehack T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
 RA Corton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
 RA Klenk H.-P., Fraser C.M., Smith H.O., Mose C.R., Venter J.C.;
 RA "Complete genome sequence of the methanogenic archaeon, Methanococcus
 jannaschii.";
 RL Science 273:1058-1073(1996).
 CC -1- SIMILARITY: STRONG TO B.SUBTILIS SPSE.
 CC -1- SIMILARITY: Contains 1 AFP-like domain.
 CC -1- SIMILARITY: Contains 1 AFP-like domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; U67549; AAB99068.1; -
 EMBL; U67549; AAB99068.1; -
 PIR; H64432; H64432.
 TIGR; M1065; -
 InterPro: IPR006014; Antifreeze_dom.
 InterPro: IPR006190; Antifreeze_like.
 InterPro: IPR004144; Neut.
 Pfam; PF01354; Antifreeze; 1.
 Pfam; PF0102; Neut; 1.
 ProSite; PS50844; AFP_LIKE; 1.
 Hypothetical protein; Complete proteome.
 DOMAIN 279 337 AFP-LIKE.
 SEQUENCE 337 AA; 37976 MW; 790CCAPF48C111B CRC64;

Query March 32.8%; Score 619; DB 1; Length 337;
 Best local similarity 40.0%; Pred. No. 8.1e-40;
 Matches 136; Conservative 65; Mismatches 115; Indels 24; Gaps 7;

QY 11 RWVGQHPCTIABIGNHQGDLDVAKMIMAKKCGADCAKPKSELEFKNRALEP 70
 DB 9 RWVGQHPCTIABIGNHQGDLDVAKMIMAKKCGADCAKPKSELEFKNRALEP 70
 QY 71 YTSKSGKTYGHHKHLFEFSHDYREIQRVAEEVGFPTASGDENAVEFLHEINVEFP 130
 DB 59 FISKYS---EYELFKSLSESEFHLKRYAKIGIMISTLDLKYVILNKMNVP 115
 QY 131 KVGSGDITNFPYLKRTAKKGRPMVSSGQSGMDTKQYQIVKPLNPF-----FLQCS 186
 DB 116 KIASGDLTPYPLLEKAKTKGPVILSTGMSDIG---EIMRAVRLVLENNGCRDITLLACIS 172
 QY 187 APFLOPDEVNLRVISEYQCLPDPPIGSGHETGIAISVAALAKVLERITLDKTKW 246
 DB 173 SYPPYSDVNLNRIKTKSLF-NIPVGRSDHTIGILAPVSVLADGVIEKHFTLDKMS 231

QY 247 GSDHSLERPELAEIVRSLVLRALGSPKPOLPCMACNEKIGKSVAKIKPECTI 306
 DB 232 GSDHSLADPEEFKXVNNRILVERKLGSGEKIPSPESHDVIVEARSLVAKRNIKKGY 291
 QY 307 LTMDELTVKVGEP-KAYPPEDIJFNLVKKVLTVEEDTI 345
 DB 292 LSVDNISFK-RPGRGIETKYSIILNRKIKNDKEEDII 329
 RESULT 3
 SPSE_BACSU STANDARD; PRT; 373 AA.
 ID SPSE_BACSU 393625;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Spore coat polysaccharide biosynthesis protein spse.
 GN SPSE OR IPA-67D OR BS037870.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_Taxid=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=98044033; PubMed=7934828;
 RA Kunz F., Ogatawara N., Moszer I., Albertini A.M., Alloni G.,
 Azevedo V., Betteiro M.G., Bessieres P., Bolotin A., Borchert S.,
 Borris R., Boursier U., Brans A., Braun M., Brignell S.C., Bron S.,
 Brouillet S., Bruschi C.V., Caldwell B., Capiano V., Carter N.M.,
 Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,
 Denton K.D., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
 Euteneier C., Fabbro J., Fabbre C., Ferrari B., Foulger D.,
 Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
 Ghim S.I., Glaeser P., Goffeau A., Gollightly E.J., Grand G.,
 Giusseppe G., Guy B.J., Haga K., Haech J., Harwood C.R., Henaut A.,
 Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,
 Joris B., Karamata D., Kasahara Y., Kjaer-Blanchard M., Klein C.,
 Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
 Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
 Lee S.M., Levine A., Liu H., Maunda S., Manel C., Medigue C.,
 Medina N., Mellado R.P., Mizuno K., Mostl D., Nakai S., Noback M.,
 Noone D., O'Reilly M., Ogawa K., Ogihara A., Oudega B., Pak S.H.,
 Pato V., Pohl T.M., Portetelle D., Potworlik S., Prescott A.M.,
 Pressac E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
 Rieger M., Rivolta C., Rocha R., Roche B., Rose M., Sadie Y.,
 Sato T., Scanlan E., Schleich S., Schroeter R., Scofield P.,
 Sekiguchi J., Sekowska A., Seror S.J., Serron P., Shin B.S., Soldo B.,
 Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
 Takeuchi M., Tanakoshi A., Tanaka T., Terstra P., Tognoni A.,
 Tosiato V., Uchiyama S., Vandenbol M., Vamler F., Vassarioli A.,
 Varsi A., Wambuit R., Wedler H., Wedler H., Weitenegger T.,
 Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 Yoshida K., Yoshikawa H.F., Zurestein E., Yoshikawa H., Danchin A.,
 RT The complete genome sequence of the Gram-positive bacterium *Bacillus*
 subtilis".
 RL Nature 350:249-256(1997).
 CC -1- PATHWAY: Spore coat polysaccharide biosynthesis.
 CC -1- SIMILARITY: Contains 1 AFP-like domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: X73124; CAAS1623.1;
 DR EMBL: 299123; CAB15613.1;
 DR PIR: S39722; S39722.
 DR Subtilisin; BGI0613; spse.
 DR InterPro: IPR006014; Antifreeze dom.
 DR InterPro: IPR006190; Antifreeze-like.
 DR InterPro: IPR004144; Neut.
 DR Pfam: PF01354; Antifreeze; 1.
 DR Pfam: PF01302; Neut; 1.
 DR PROSITE: PS50844; AFP_LIKE; 1.
 KW Complete proteome.
 FT DOMAIN 305
 FT SEQUENCE 373 AA; 40889 MW; 3CFBDFACE0DCE38 CRC64;
 Query Match 19.6%; Score 370; DB 1; Length 373;
 Best Local Similarity 30.1%; Freq. No. 7.8e-21;
 Matches 107; Conservative 55; Mismatches 167; Indels 26; Gaps 9;
 QY 13 VGGQPCFIARIGGNHGGDLDVAKMIRMAKCGADCAKFKQSLERFKALKERPT 72
 DB 11 VKXAPVITIAAGINHGKLDQALIDAAAEAGADAVKQM-----FQADMVYKDPOL 66
 QY 73 SRSWGTGTEHK-RHLEFSDQYRELQYAEVGIFFTSAGDEMAVEFHELVNPF 130
 DB 67 YTTAGKQVSIPISTVQSMEMPALWILPLDYCREKQVIFLSTVCDEGSDLLQSTSPSAF 126
 QY 131 KYGSDTNPFPLETKATKGRPMVYSSGQMSMDCKQYQYIYKPL-NPNPCTLOCTSAVP 189
 DB 127 KLAETINHLPLKTVARNRPMPTSTMGAEISDVHAKRTIRACGNQOIAHNCVAKYP 186
 QY 190 LQPEDVNLVISEYOKLPPDIPIGYSG-ETGIAISVAVALGAKVLERHITLDTKWS 248
 DB 187 APPEVSNLSVIMLAAPPEAVIGSDSHSEHTAPCAAVRGALIEHFTIDNLPQA 246
 QY 249 DHSASLEPELAEIVRSR-----LVRAAGSPKPOLPCMACNEKIGKSV 295
 DB 247 DHSFALNPDELKEMWDGKRKEAEIKQGITKPVSEKLSSTYKTTALBGRNFAVKRI 306
 QY 296 VAKVPIBEGITLTMULTY-KVGE-PKAYPPEDIJFNLV-GKVLVTVZEEDTI 346
 DB 307 FTTAIQKGEARSBNINAVLRGCKPQGGHPR-FHELTGVRAYADIPADTGIV 360
 RESULT 4
 SMIB_MOUSE STANDARD; PRT; 1248 AA.
 ID SMIB_MOUSE 0920F6;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Structural maintenance of chromosomes 1-like 2 protein (SMC1beta
 DE protein).
 GN SMC1L2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_Taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A., PARTIAL SEQUENCE, FUNCTION, CHARACTERIZATION, AND
 RP INTERACTION WITH SMC3.
 RC TISSUE-Testis;
 RX MEDLINE=21448993; PubMed=11564881;
 RA Revankova Z., Elpe M., Heyling C., Gross B., Jessberger R.;
 RT "Novel meiosis-specific isoform of mammalian SMC1".
 RL Mol. Cell. Biol. 21:6984-6998(2001).
 CC -1- FUNCTION: Meiosis specific component of cohesin complex. The
 CC cohesin complex is required for the cohesion of sister chromatids
 CC after DNA replication. The cohesin complex apparently forms a

large proteinaceous ring within which sister chromatids can be trapped. At anaphase, the complex is cleaved and dissociates from chromatin, allowing sister chromatids to segregate. The meiosis-specific cohesin complex probably replaces mitosis specific cohesin complex when it dissociates from chromatin during prophase I.

-1- SUBUNIT: Forms a heterodimer with SMC3. Component of a meiosis-specific cohesin complex, probably composed of the SMC2L2 and SMC3 heterodimer attached via their hinge domain, RAD21 (or its meiosis-specific related protein REC8), which link them, and STAG3, which interacts with RAD21 or REC8.

-1- SUBCELLULAR LOCATION: Nuclear protein. Associates with chromatin. In prophase I stage of meiosis, it is found along the axial elements of synaptonemal complexes. In late-pachyene-diplophase, the bulk of protein dissociates from the chromosome arms probably because of phosphorylation by PLX, except at centromeres, where cohesin complexes remain. It however remains chromatin associated at the centromeres up to metaphase II. At anaphase II, it dissociates from centromeres, allowing chromosome segregation.

-1- TISSUE SPECIFICITY: Testis and ovary specific. Not expressed in somatic cells.

-1- DOMAIN: The flexible hinge domain, which separates the large intramolecular coiled coil regions, allows the heterotypic interaction with the corresponding domain of SMC3, forming a V-shaped heterodimer. The two heads of the heterodimer are then connected by different ends of the cleavable RAD21 or REC8 protein, forming a ring structure (By similarity).

-1- SIMILARITY: Belongs to the SMC family. SMC1 subfamily.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: AF303827; MAF09333.1; --

MGD: MGI:2154049; SMC1L2.

DR GO: GO:0000775; C:chromosome, pericentric region; IDA.

DR GO: GO:0000795; C:synaptonemal complex; IDA.

DR GO: GO:0005677; F:DNA binding; IDA.

DR GO: GO:0005615; F:protein binding; IPT.

DR GO: GO:0007062; P:sister chromatid cohesion; IDA.

DR InterPro: IPR003439; ABC transporter.

DR InterPro: IPR003405; SMC_C.

DR InterPro: IPR003395; SMC_N.

DR Pfam: PF02483; SMC_C; 1.

DR Pfam: PF02463; SMC_N; 1.

KM Meiosis; Cell cycle; Chromosome partition; ATP-binding; Coiled coil; Nuclear protein.

KM NP BIND 32 ATP (POTENTIAL).

FT DOMAIN 163 502 COILED COIL (POTENTIAL).

FT DOMAIN 503 665 FLEXIBLE HINGE.

FT DOMAIN 666 912 COILED COIL (POTENTIAL).

FT DOMAIN 1123 1158 ALA/ASP-RICH (DA-BOX).

FT DOMAIN 1248 144513 MW: 6C310B46217BC94D CRC64;

SEQUENCE

Query Match 6.0%; Score 113.5; DB 1; Length 1248;

Best local similarity 23.5%; Pred. No. 1.1;

Matches 88; Conservative 56; Mismatches 138; Indels 93; Gaps 19;

DB 31 GDLVAKKIRMAKCGADCAFCQSELEFKRNR--ALERTYKSH--MKRTYGHNR 86

DB 166 GEYFAKKKLOKRAE-----DAQFHVAKKVAER--KAKIKKEEAHQ 210

DB 87 H-----LESGDYRELFQRAEEVGIFPTASGMDVAEFLHEINVPFVAGSDT--NMP 141

DB 211 NLLEIKIKIKIQLMKFLQYVNEKIVLVNTELEQM-----DGNLSVAV--DTLSHR 260

DB 142 YLETKAKGKPMYISSGMSMDTKKQYQIVQFLANFGLQCTSAIYPLQPEDVNLKVS 201

DB 261 NIFKAKKQYGMLTRQLOQTAKELKSEVALINDKRPQYTKAVENTSHHLKLDLSEKLT 320

202 EYQKLPDIPIGYSGHETGIAISVAVALCAKVLKRNHTIDPKTKGSDHSAS---LEPGE 258

DB 321 DNEK-----QSKQDQDGRALVAEIA-----DDRWKSPFKQMEKILQKGR 363

QY 259 LAELVRS-----VALVRALQSPFKQLPCC---EMAGNEKLG-----KS 294

DB 364 DIELHNSQLRYRLKQVARKVQIMTQQLQKQWPKAKERKRLAEPKRHSPTQGNLKQ 423

QY 295 VAVRV-----KIPGTLTMDMTLVKVGEPKAPPPEDIPNLVSKKVLVTWEEDTITM 346

DB 424 IKQDIEHKRIEKLSEYVTCMDCLQEDKXQGBALKE--IENYKSMSEVN--EELSILR 481

QY 347 EEV-----VDNH--GRX 356

DB 482 NELQNAQIDNHGKGR 496

RESULT 5

ANP3 RHIDE STANDARD; PRT; 134 AA.

AC P35753;

DT 01-JUN-1994 (Rel. 29, Created)

DT 01-JUN-1994 (Rel. 29, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Antifreeze peptide RD3.

OS Rhigophylla dearborni (Antarctic eel pout) (Lyocodichthys dearborni).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; OC Acanthomorphi; Acanthopterygii; Percormorpha; Perciformes; Zoarcoidei; OC Zoarcidae; Lycodichthys.

NCBI_TaxID=8201;

OX [1]

RP SEQUENCE.

RP MEDLINE=95210308; PubMed=7696304;

RA Wang X., Davies A.L., Cheng C.-H.C.;

RT "Antifreeze peptide heterogeneity in an antarctic eel pout includes an unusually large major variant comprised of two 7 kDa type III AFPs linked in tandem.";

RL Biochim. Biophys. Acta 1247:163-172(1995).

RN [2]

RP STRUCTURE BY NMR OF 1-73.

RX MEDLINE=99355969; PubMed=10423534;

RA Miura K., Ogiya S., Hoshino T., Nemoto N., Odaire M., Nitta K., Tsuda S.;

RT "Determination of the solution structure of the N-domain plus linker of antarctic eel pout antifreeze protein RD3.";

RL J. Biochem. 126:387-394(1999).

CC -1- FUNCTION: Antifreeze proteins lower the blood freezing point by absorbing ice and inhibiting its growth.

CC -1- SIMILARITY: Belongs to the type-III AFP family.

CC -1- SIMILARITY: Contains 2 AFP-like domains.

CC PIR: S53514; S53514.

DR PDB: 3NLA; 23-FEB-99.

DR PDB: 3RDN; 23-FEB-99.

DR PDB: 1C89; 28-FEB-01.

DR PDB: 1C8A; 28-FEB-01.

DR InterPro: IPR006014; Antifreeze dom.

DR InterPro: IPR006190; Antifreeze-like.

DR InterPro: IPR006013; AntifreezeIII.

DR Pfam: PF01354; Antifreeze; 2.

DR PRINTS: PR00357; ANTIFREEZEIII.

DR ProDom: PD003258; AntifreezeIII; 2.

DR PROSITE: PS00844; AFP_LIKE; 2.

KM Antifreeze protein; Multigene family; Repeat; 3D-structure.

KM DOMAIN 4 63 AFP-LIKE 1.

FT DOMAIN 65 70 LINDER.

FT DOMAIN 74 133 AFP-LIKE 2.

FT SITE 9 9 IMPORTANT FOR ICE-BINDING (BY SIMILARITY).

FT SITE 14 14 IMPORTANT FOR ICE-BINDING (BY SIMILARITY).

FT SITE 18 18 IMPORTANT FOR ICE-BINDING

FT SITE 44 44 (BY SIMILARITY)
 FT SITE 79 79 (BY SIMILARITY)
 FT SITE 84 84 (BY SIMILARITY)
 FT SITE 88 88 (BY SIMILARITY)
 FT SITE 114 114 (BY SIMILARITY)
 FT STRAND 4 7
 FT STRAND 11 11
 FT STRAND 19 21
 FT STRAND 22 25
 FT STRAND 33 33
 FT STRAND 37 40
 FT STRAND 41 42
 FT STRAND 45 45
 FT STRAND 49 49
 FT STRAND 51 52
 FT STRAND 53 54
 FT STRAND 57 59
 FT STRAND 60 60
 SQ SEQUENCE 134 AA; 14480 MW; F7P208BF3E2CA54 CR64;
 Query Match 5.8%; Score 110; DB 1; Length 134;
 Best Local Similarity 37.4%; Pred. No. 0.13; Mismatches 12; Indels 8; Gaps 2;
 Matches 34; Conservative 12; Mismatches 37; Indels 8; Gaps 2;
 QY 269 VERALGSPFKQLPCEVACNCKLQ-----KSYVAKKREPEGLTMDLTVKVGPKAY 322
 DB 45 VNRAYV-PLGTLTMDPKVKNYEDGTTSPKSYVANGELINLTLTVKAKAEVSPKGI 102
 QY 323 PREDIFNLVKKYLVYVEEDDTMEELVDNH 353
 DB 103 PSEISKLWGVNRAVYLDQTLMDPKVKNY 133
 RESULT 6
 PARE_MYCGE STANDARD; PRT; 633 AA.
 AC P47445;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Topoisomerase IV subunit B (EC 5.99.1.-).
 GN PARE OR MG203.
 OS Mycoplasma genitalium.
 OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
 OX NCBI_TaxID=2097;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 33530 / G-37;
 RX MEDLINE=56026346; PubMed=7569993;
 RA Fraser C.X., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
 RA Fleischman R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
 RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhmann J.L.,
 RA Nguyen D.T., Uterback T.R., Sauder D.M., Phillips C.A., Merrick J.M.,
 RA Tomb J.-P., Dougherty B.A., Bort K.F., Hu P.-C., Luster T.S.,
 RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.:
 RT "The minimal gene complement of Mycoplasma genitalium.",
 RL Science 270:397-403(1995).
 RP SEQUENCE OF 406-633 FROM N.A.
 RC STRAIN=ATCC 33530 / G-37;
 RA Bailey C.C., Younkins R., Huang W.M., Bott K.F.;
 RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: Topoisomerase IV is essential for chromosome
 CC segregation. It has relaxation events required during the replication
 CC of a circular DNA molecule (by similarity).
 CC -1- SUBUNIT: Composed of two subunits: parC and parE.

CC -1- SIMILARITY: Belongs to the type II topoisomerase family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: U39700; AAC71421.1;
 DR EMBL: U25549; AAC33990.1;
 DR PIR: D64222; D64222.
 DR HSP: P06982; LAJ6.
 DR TIGR: MG203;
 DR InterPro: IPR003594; ATPbind_ATPase.
 DR InterPro: IPR002288; DNA_gyraseB_C.
 DR InterPro: IPR001241; DNA_topoisom.
 DR InterPro: IPR005740; TopoIV_B_gpos.
 DR InterPro: IPR006171; Toprim.dom.
 DR Pfam: PR00204; DNA_gyraseB_1.
 DR Pfam: PR00986; DNA_gyraseB_C_1.
 DR Pfam: PR02518; HATPase_C_1.
 DR Pfam: PF01751; Toprim_1.
 DR PRINTS: PR00418; TP12FAMILY.
 DR ProDom: PD149633; DNA_gyraseB_C_1.
 DR SMART: SM00387; HATPase_C_1.
 DR SMART: SM00433; TOP2c_1.
 DR TIGRPFAM: TIGR01058; PARE_gpos_1.
 DR PROSITE: PS00177; TOPOISOMERASE_II; 1.
 KW Isomerase; Topoisomerase; ATP-binding; Complete proteome.
 SQ SEQUENCE 633 AA; 71569 MW; 249B854169A8CB6 CR64;

Query Match 5.7%; Score 108.5; DB 1; Length 633;
 Best Local Similarity 18.1%; Pred. No. 1.1; Mismatches 139; Indels 147; Gaps 18;
 Matches 77; Conservative 63; Mismatches 139; Indels 147; Gaps 18;
 QY 26 GONHGDLDVANKRMTRMEKCG-----ADCAFTQSS-----ELEPK 61
 DB 137 GOIYGVYVAGGKIIQAKKIDGTTSGTTVSPFHADPKVFKKAQPSNIIKRLKELSTL 196
 QY 62 FPKKLEREYTSKHSNGKTYGKHLBESHQYRELQYASEVGIFPTASGDMANAVEF 121
 DB 197 FAKLKL-----TFTDQK-----TKITYFFETSGL-----VGF 224
 QY 122 LRLNVPFFKVGSGDTNPPYLEKTAKGR-----PWYSGQSGMDTKQYQIVKEL 175
 DB 225 LDEINNTVETLG-----QKTLIKGKDGIEVEVVFQFNQSDQETILSPANSIKTF 274
 QY 176 -----NPNRCFL-----QCTSAVPLQPEVDNLAVTSEYKL-----PFDIPGYSG 216
 DB 275 EGGSHENGFLAISVDVNSYCKRNLLKEDKNFQUSEIRQGLNALIKTNLBERKNIAPFG 334
 QY 217 HETGIAISVAVALGAKVLERH-----ITLDKTKNGSDHSASLPGLAEIV 263
 DB 335 QIKSKLFSKEVNVNVAIVQCHVFOFLERNNDAXLIIDL-----LNAKIKEDQI 385
 QY 264 KSVRLVERALSGPTKO-----LIPGE-----MACNEKLGSVAKVXI 301
 DB 386 KQQRLEKKSLSPPQEKILFGLAPQCYTKSEKELFIVEGDSAGSTAVKGRIRIQALI 445
 QY 302 P-EGTILTMDLTVKVGPKAYPEDIENL-----VGKVLVYVEEDDTIMEELV-----DN 352
 DB 446 PLRGKLVANVEKLNK-----KEALTNEILLVILICIGILTNINIKLKYKGIIMTDAON 502
 QY 353 ECKKIK 358
 DB 503 DGAHIQ 508
 RESULT 7
 ANP2_AUSBR STANDARD; PRT; 63 AA.
 ID ANP2_AUSBR

Db 334 G 334

RESULT 9

DB GATB_AQUAE STANDARD; PRT; 478 AA.
AC 066766;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE AspArgyl/glutamyl-tRNA(Asn/Gln) amidotransferase subunit B
DE (EC 6.3.5.-) (asp/glu-ADT subunit B).
GN GATB OR AQ_461.
OS Aquifex aeolicus.
OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
OX NCBI_Taxid=63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF5;
RX MEDLINE=98196666; PubMed=9537320;
RA Decker G., Warren P.V., Gaasterland T., Young W.G., Lemox A.L.,
RA Graham D.E., Overbeek R., Sread M.A., Keller M., Aubay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.D., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus";
RL Nature 392:353-358(1998).
CC -1- FUNCTION: Allows the formation of correctly charged Asn-tRNA(Asn)
CC or Gln-tRNA(Gln) through the transamidation of misacylated Asp-
CC tRNA(Asn) or Glu-tRNA(Gln) in organisms which lack either or both
CC of asparaginyl-tRNA or glutamyl-tRNA synthetases. The reaction
CC takes place in the presence of glutamine and ATP through an
CC activated phospho-Asp-tRNA(Asn) or phospho-Glu-tRNA(Gln) (by
CC similarity).
CC -1- CATALYTIC ACTIVITY: ATP + L-glutamyl-tRNA(Gln) + L-glutamine = ADP
CC + phosphate + L-glutamyl-tRNA(Gln) + L-glutamate.
CC -1- CATALYTIC ACTIVITY: ATP + L-asparaginyl-tRNA(Asn) + L-glutamine = ADP
CC + phosphate + L-asparaginyl-tRNA(Asn) + L-glutamate.
CC -1- SIMILARITY: Heterotrimer of A, B and C subunits (By similarity).
CC -1- SIMILARITY: Belongs to the gatB/gatB family. GatB subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; AB000691; AAC06727.1; -
DR PIR; B70342; B70342.
DR HAMAP; MF 00121; - 1.
DR InterPro; IPR000443; GatB.
DR InterPro; IPR006107; GatB_cent.
DR InterPro; IPR006075; GatB_N.
DR InterPro; IPR003789; GatB_Ygey.
DR Pfam; PF01162; GatB; 1.
DR Pfam; PF02934; GatB_N; 1.
DR Pfam; PF02637; GatB_Ygey; 1.
DR TIGRfam; TIGR00133; GatB_N.
DR PROSITE; PS01234; GatB; 1.
KW Protein biosynthesis; Ligase; Complete proteome.
SQ SSQUNCE 478 AA; 55040 MW; F725AE78944BD79A CRC64;

Query Match 5.4%; Score 106.5; DB 1; Length 478;
Best local Similarity 19.6%; Pred. No. 1.1; Indels 101; Gaps 14;
Matches 62; Conservative 55; Mismatches 100; Indels 101; Gaps 14;

QY 93 DQRELORYAEVGFITFASGMDNAVEFHELANVFPFKVSGD-----TNNPPL 143
DB 171 EKLRLNRYA-----SKADWEKQALCDINVISIRPKSKEFSTREIVEIKVNSFRFV 223
QY 144 ETKAKKGRPVNTSSGMSQDPMKOVYQIVKPLNPNCFLOCTSAFPLQ----- 191

Db 224 QYALE-----YETEROINVEBEGEVQETRTFDP-----QTKTTPMRTKEAEADRYRP 274
QY 192 -PEDVNLRY-----ISEYKLFEPDIP-----IGYSGHETGIAISVAVALGAKYLE 236
Db 275 DEDLVPLAKKKEWIMEIKKNPHELPDQFERLIKGYGSEYAGILVNHKEVGDFFBEAV 334
QY 237 REITLTK-----TW-----KG-SHNSASLAEGLAELVRSRLVERALGSPKQLL 281
Db 335 RHFKEPKGIWMLINDLLGLRDKGISIEESPVEKHEHAEVLKLIK----- 380
QY 282 PCMAKNEKLGKSVYAKVIFPGTILTMQMLTVKVGEPYAPPEDIFNLWKGVLTVEE 341
Db 381 --EKVISTKIGEVKEM-----VERCK-----TPSGIVEKGIKQITDENQ 420
QY 342 DDTIMEELVNDGKKIK 358
Db 421 IKELVKIFKEKPKKEVE 437

RESULT 10

DB MYCPN STANDARD; PRT; 1140 AA.
AC P75405;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein M5N376 (A19_071140).
GN M5N376 OR M5460.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_Taxid=2104;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE=97103885; PubMed=8948633;
RA Himelreich R., Hilbert H., Plagens H., Pirkl E., Li B.-C.,
RA Herrmann R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT pneumoniae";
RL Nucleic Acids Res. 24:4420-4449(1996).
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; STRONG, TO M.PNEUMONIAE
CC M5N375.
CC -1- SIMILARITY: IN THE CENTRAL SECTION; STRONG, TO M.PNEUMONIAE
CC M5N374.
CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; STRONG, TO M.PNEUMONIAE
CC M5N373.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; AE000045; AAB96108.1; -
DR PIR; S73786; S73786.
DR Hypothetical protein; Transmembrane; Complete proteome.
KW TRANSMEM 8 28
FT TRANSMEM 1098 1118 POTENTIAL.
SQ SEQUENCE 1140 AA; 130383 MW; 8FA9406C57D8886 CRC64;

Query Match 5.4%; Score 102.5; DB 1; Length 1140;
Best local Similarity 21.3%; Pred. No. 6.7; Indels 83; Gaps 16;
Matches 84; Conservative 60; Mismatches 167; Indels 83; Gaps 16;

QY 22 IABIGONHGGDIDVAKRMIRNA-----KCGADCAKFOK--SELKFKRKALERY 71
Db 598 VSAIGFQDEGGIRDSFINKVALTPKSSKTKLANDDASLEAAAEISYKH-----YRQN 653
QY 72 TSKHSGKTYGGEKHLSESHDQRELORYAEVGIFFTASGMD--EAAVEFLHNLNPF 129
Db 654 FROASWMDXNSQTKSIVOSTDINDERFERFORDINNYLKVQGISFTEIKVNAVHRVDA-M 712

Qy 130 FKVSGDTNNPPEYKTKKGRPMVYISGQMSMDTKOVQIYVPLNPN----- 178
 Db 713 LNAKSDPKLASVOSTANK-----YGLNLSNPPYVGQFYVAVVYVYVANDJANGORANNA 767
 Qy 179 ---FCPLQ-----CISAVPLQEDVNLRLVISEYQKLPEDIPIGYSGHGTGIAISVAA--- 227
 Db 768 KSYFYIIGLDKQKQSSVLYRFE--NKOKLYSLESLANDSKGLYKVNKXKAIIGAKQNO 825
 Qy 228 -----VALGAKVLERHITL-----DKTWKSDHSASLEPGELELYRSVRL--YE 270
 Db 826 NLYLDTNMNAALKNLTNAELTFTASADNSAKLSTENARNDBGFLSENVSGLIAYVE 885
 Qy 271 PALG-----SPTKOLPEEMAKONEKLGKSVAVAKRIPBGTILTMDEM-----LT 313
 Db 886 RMTGKKLPLKERVSNKEDKNLKLRLNSFTLDDKKKLYKQDP--SYINQIVBEAKQYN 943
 Qy 314 VKVGSPPKAYPPEDIPNEVGKVLVTVVEEDDTIME 347
 Db 944 VLVSEKGDPPESDKNIF-KTILTNPEQSTVIK 976

RESULT 11

ANP1_AUSBR STANDARD; PRT; 63 AA.
 ID P12100;
 AC 01-OCT-1989 (Rel. 12, Created)
 DT 01-OCT-1989 (Rel. 12, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DB Antifreeze peptide ABL.
 OS Austrolychthys brachycephalus (Antarctic eel pout).
 CC Austrolychthys brachycephalus (Antarctic eel pout).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Neuteleostomi;
 CC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neteleostei;
 CC Acanthomorpha; Acanthopterygii; Percormorpha; Perciformes; Zoarcoidel;
 CC Zoarcidae; Austrolychthys.
 CC NCBI_Taxid=8195;
 RN (1)
 RP SEQUENCE.
 RX MEDLINE=89323219; PubMed=2752054;
 RA Cheng C.-H.C.; Devries A.L.;
 RT "Structures of antifreeze peptides from the antarctic eel pout,
 RL Austrolychthys brachycephalus."
 RL Biochim. Biophys. Acta 987:55-64(1989).
 CC -1- FUNCTION: Antifreeze proteins lower the blood freezing point by
 CC absorbing ice and inhibiting its growth.
 CC -1- SIMILARITY: Belongs to the type-III AFP family.
 CC -1- SIMILARITY: Contains 1 AFP-like domain.
 CC PIR: S04973; S04973.
 DR HSSP: P35753; 3RD.
 DR InterPro: IPR006014; Antifreeze_dom.
 DR InterPro: IPR006190; Antifreeze_like.
 DR InterPro: IPR006013; AntifreezeIi.
 DR Pfam: PF01354; Antifreeze; 1.
 DR PRINTS: PR00357; ANTIFREEZEIII.
 DR PRODOM: PD003258; AntifreezeIi; 1.
 DR PROSITE: PS00644; AFP_LINE; 1.
 KM Antifreeze protein.
 FT SITE 3 62 APP-LIKE.
 FT DOMAIN 3 8 IMPORTANT FOR ICE-BINDING
 FT SITE 8 8 (BY SIMILARITY).
 FT SITE 13 13 IMPORTANT FOR ICE-BINDING
 FT SITE 13 13 (BY SIMILARITY).
 FT SITE 17 17 IMPORTANT FOR ICE-BINDING
 FT SITE 17 17 (BY SIMILARITY).
 FT SITE 43 43 IMPORTANT FOR ICE-BINDING
 FT SITE 43 43 (BY SIMILARITY).
 SQ SEQUENCE 63 AA; 6846 MW; AC84FD14247153B4 CRC64;
 Query Match 5.4%; Score 102; DB 1; Length 63;
 Best Local Similarity 39.3%; Pred. No. 0.2;
 Matches 24; Conservative 10; Mismatches 27; Indels 0; Gaps 0;
 Qy 293 KSVAVAKVIRPGTILLTMKLTVMKVGPKAYPPEDIPNEVGKVLVTVVEEDDTIMEELVND 352

Db 2 KSVVASGLTIRINLTPAMKAKEVSPKGIPEBESKIVQMQRNRAVNUDETILAPDMVKY 61
 Qy 353 H 353
 Db 62 Y 62

RESULT 12

MYR_RAT STANDARD; PRT; 1107 AA.
 ID MYR_RAT
 AC Q6356;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DB Myosin Ie (Myosin heavy chain myr 3).
 GN MYOIE OR MYR3.
 OS Rattus norvegicus (Rat).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 CC NCBI_Taxid=10116;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=Sprague-Dawley;
 RX MEDLINE=95247829; PubMed=7730414;
 RA Stoeckler H.B.; Ruppert C.; Reinhard C.; Bahler M.;
 RT "A novel mammalian myosin I from rat with an SH3 domain localizes to
 RL Con A-inducible, F-actin-rich structures at cell-cell contacts."
 RL J. Cell Biol. 129:819-830(1995).
 CC -1- FUNCTION: Myosins are actin-based motor molecules with ATPase
 CC activity. Unconventional myosins serve in intracellular movements.
 CC Their highly divergent tails are presumed to bind to membranes
 CC compartments, which would be moved relative to actin filaments (By
 CC similarity).
 CC -1- SIMILARITY: BELONGS TO THE MYOSIN HEAVY CHAIN FAMILY.
 CC -1- SIMILARITY: Contains 1 IQ domain.
 CC -1- SIMILARITY: Contains 1 SH3 domain.
 CC -----
 CC EMBL: X74815; CAA52815.1; --
 DR PIR: S52517; S52517.
 DR HSSP: P08799; 1KND.
 DR InterPro: IPR000048; IQ_region.
 DR InterPro: IPR001609; myosin_head.
 DR InterPro: IPR001452; SH3.
 DR Pfam: PF00612; IQ_1.
 DR Pfam: PF00063; myosin_head; 1.
 DR Pfam: PF00018; SH3; 1.
 DR PRINTS: PR00193; MYOSINHEAVY.
 DR PRINTS: PR00452; SH3DOMAIN.
 DR PRODOM: PD000355; myosin_head; 1.
 DR PRODOM: PD000066; SH3; 1.
 DR SMART: SM00242; MYSC; 1.
 DR SMART: SM00326; SH3; 1.
 DR PROSITE: PS50096; IQ_1.
 DR PROSITE: PS50002; SH3; 1.
 DR Myosin; ATP-binding; Actin-binding; Calmodulin-binding; SH3 domain;
 KM Multigene family.
 FT DOMAIN 1 679 MYOSIN HEAD-LIKE.
 FT DOMAIN 1 724 IQ.
 FT DOMAIN 1050 1107 SH3.
 FT DOMAIN 581 591 ACTIN-BINDING (POTENTIAL).
 FT NP BIND 112 119 AFP (POTENTIAL).
 SQ SEQUENCE 1107 AA; 126826 MW; B9DBFBCE047148 CRC64;
 Query Match 5.4%; Score 101; DB 1; Length 1107;

Best Local Similarity 22.2%; Pred. No. 8.4;
Matches 70; Conservative 43; Mismatches 114; Indels 88; Gaps 15;

QY 47 GADCAKPKOKSELEF---KENPKALERTPTSKSKSKTYGKRRHLEPSHQYRELQRYA 102
DB 476 GADQCLLOKLOQMOISHEHFN-----SWMOGIIHRYAGKSYDMDGFCENNR 523
QY 103 BEVGJFFTFASGMDENAVFELHNPFPFVSGDTPNNFYELKTAAGKGPVWISSGQSM 162
DB 524 D---VLF---MD---LHIMOSELPFK-----SIFENNQANXKXGPTTAGSKIK-- 566
QY 163 DTMKQVOYQVKEPLNPFCLQCTSAV-----FLQPDVNLRYVSEYQKPPDIP 211
DB 567 ---KQANDEAVSTL-----MKCTPHYRCIKENETKPKPMWESRSYKQVETLQKENIR 617
QY 212 IGVSGHEHICIAISVAVALGAVLEBHITLTK-----TWKSGDHSASLEPGELEHVSVR 267
DB 618 VRRAGT-----AYRRVQKFLQRYALITKATPFWGDKQVYL-----HILQSYN 663
QY 268 IYERALGSPFKQPLPCENACNEKLGKSVYAKVKEGIIITMDMLTWKGEKAVPPEDI 327
DB 664 M-----DSQGF-----QLGSKVVF-IPAPSLFLEEMERKDYGVAFVIOKTW 706
QY 328 FNLVNGKVLVTVYED 342
DB 707 RKFVARKKYVQWED 721

RESULT 13

BD11_BACSU STANDARD; PRT; 469 AA.
ID BGLI_BACSU
AC F40740; 032287;
DT 01-FEB-1995 (Rel. 31, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE beta-glucosidase (EC 3.2.1.21) (Gentilibase) (Cellubiose) (Beta-D-glucoside glucosylhydrolase) (Amygdalase).
GN BGLH OR N17D OR BSU39260.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
CX NCBI_Taxid=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / Marburg;
RX MEDLINE=95189730; PubMed=7883710;
RA Le Coq D., Lindner C., Krueger S., Steinmetz M., Stuelke J.,
RT "New beta-glucoside (bg1) genes in Bacillus subtilis: the bg1 gene
RT product has both transport and regulatory functions similar to those
RT of BglP, its Escherichia coli homolog.";
RL J. Bacteriol. 177:1527-1535(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / BSC21A1;
RX MEDLINE=95213088; PubMed=7704263;
RA Yoshida K.-I., Sano H., Seki S., Oda M., Fujimura M., Fujita Y.,
RT "Cloning and sequencing of a 29 kb region of the Bacillus subtilis
RT genome containing the hut and wpa loci.";
RL Microbiology 141:337-343(1995).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=9804033; PubMed=9384377;
RA Kunst P., Ogasawara N., Mosser I., Albertini A.M., Alloni G.,
RA Azevedo V., Bacterio M.G., Bessieres P., Bolotin A.R., Borchert S.,
RA Borriss R., Boursier L., Brans A., Brann M., Brignell S.C., Bron S.,
RA Bruller S., Brusch C.V., Caldwell B., Capiano V., Carter N.M.,
RA Choi S.K., Codani J.U., Conerton I.F., Cummings N.J., Daniel R.A.,
RA Dautot F., Devine K.M., Dusterhoft A., Enrich S.D., Emerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Funa S., Galizzi A., Galleron N.,
RA Grim S.Y., Glaeser P., Goffeau A., Golightly E.J., Grand G.,
RA Griespiel G., Guy B.J., Haga K., Haelech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hulle M.F., Itaya M., Jones L.,

RA Joris B., Karamata D., Kaashara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapides A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Nodack M.,
RA Noone D., O'Reilly M., Ogawa K., Ogilwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portelle D., Portelle S., Prescott A.K.,
RA Priesecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,
RA Sato T., Scanlan E., Schleich S., Schreuter R., Scoffone P.,
RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpetra P., Tognoni A.,
RA Toesato V., Uchiyama S., Vandendol M., Vannier F., Vassartoli A.,
RA Viati A., Wambut R., Wedler E., Wedler H., Weltzenegger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Zanchin A.,
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis".
RL Nature 390:249-256(1997).
CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing beta-D-
CC glucose residues with release of beta-D-glucose.
CC -1- SIMILARITY: Belongs to family 1 of glycosyl hydrolases.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: Z34526; CAA84287.1; -
CC EMBL: D18156; BAA06653.1; -
CC EMBL: D29985; BAA06657.1; -
CC EMBL: Z99124; CAB15962.1; ALT_INIT.
CC PIR: H69593; H69593.
CC HSSP: P11546; 1PBG.
CC Subtilisin; BGI0935; bg1H.
CC InterPro: IPR001360; Glyco_hydro_1.
CC Pfam: PF00232; Glyco_hydro_1; 1.
CC PRINTS: PR00131; GLHYDRASE1.
CC ProDom: PD00650; Glyco_hydro_1; 1.
CC PROSITE: PS00572; GLYCOSYL_HYDROL_F1_1; 1.
CC PROSITE: PS00553; GLYCOSYL_HYDROL_F1_2; 1.
CC KW Hydrolase; Glycosidase; Cellulose degradation; Complete proteome.
CC ACT_SITE 175 175 PROTON DONOR (POTENTIAL).
CC ACT_SITE 368 368 NUCLEOPHILE (BY SIMILARITY).
CC ACT_SITE 402 402 A -> R (IN REF. 1).
CC CONFLICT 402 402
CC FT
CC FT
CC SEQUENCE 469 AA; 53289 MW; 4CCCA25F64C1C1C1 CRC64;

Query Match

5.3%; Score 100.5; DB 1; Length 469;

Best Local Similarity 22.2%; Pred. No. 3.2;
Matches 72; Conservative 50; Mismatches 82; Indels 121; Gaps 20;

QY 79 KTYG-EKRRHLEPSHQYR-LQRYAEYGFPTASGDMBAVFLHELV-----PEF 130
DB 138 KNYGWMKRYIEF-YERYAKTVFRYQHKYVMT-----FNEINVLHAP- 184
QY 131 KVGSGDTNFPYLKTAAGKGPVWISSGQSMDFYKQV--YQIVK-----PLN 176
DB 185 -TGGS-----LVREGEKTLNMYQAHQGFVASALAVKAGHDIIIDS 226
QY 177 FNFCLQCTSAVPL--QEDVNLRYVISEYQKL-PPDIP1-GYSGH-----ETGIALSV 225
DB 227 KIGCIATNTYTPMTSKEDVYFAMENNERKTLFESDVGARQAYPGWRYALNNIIZEM 286
QY 226 AAVNAGAVLEBRHTLDTWVGSDH-----SASLEGEALVRSVRLVERALGSPKQL 281
DB 287 AS--GDELLHEHVD--YIGFSTYMSALSTDEELA----- 320
QY 282 PCENACNEKLGKSVYAKYK-----IPGTLITMDMLTWKGEPRVAPPEDI 328

Thu May 13 11:53:10 2004

us-09-930-440b-6.rsp

Page 10

D5 321 -----KSGNLLGVKYNPKLSSNGWQIDPKGRLITLTYRQKP-LPIVENGL 371
QY 329 NLVGRKVLVTEEDPTIMEELVDNH 353
DB 372 GAVDK-----VBEODTODDPRINY 391

RESULT 14

MYO_HUMAN STANDARD; PRT; 1109 AA.
ID MYO_HUMAN
AC Q12965;
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Myosin Ic (Myosin Ic).
GN MYOIC OR MYOIC.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OK NCBI_Taxid=9606;
RN (1)
RP SEQUENCE OF 1-245 FROM N.A.
RX MEDLINE=95018277; PubMed=7932763;
RA Bement W.M., Birch J.A., Mooseker M.S.;
RT "Cloning and mRNA expression of human unconventional myosin-IC. A
RT homologue of amoeboid myosins-I with a single IQ motif and an SH3
RT domain";
RT J. Mol. Biol. 243:356-363(1994).
CC -1- FUNCTION: Myosins are actin-based motor molecules with ATPase
CC activity. Unconventional myosins serve in intracellular movements.
CC Their highly divergent tails are presumed to bind to membranous
CC compartments, which would be moved relative to actin filaments (By
CC similarity).
CC -1- SIMILARITY: BELONGS TO THE MYOSIN HEAVY CHAIN FAMILY.
CC -1- SIMILARITY: Contains 1 IQ domain.
CC -1- SIMILARITY: Contains 1 SH3 domain.
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: U14391; AAA62667.1; -
CC PIR: S53601; S53601.
CC HSPF: P08799; IAMD.
CC GeneW: HGNC:7599; MYOIE.
CC MIM: 601479; -
CC DR GO: GO:0015629; C:actin cytoskeleton; TAS.
CC DR GO: GO:0003774; F:motor activity; TAS.
CC DR GO: GO:0006570; F:myosin ATPase activity; TAS.
CC DR InterPro: IPR000048; IQ region.
CC DR InterPro: IPR001608; myosin_head.
CC DR InterPro: IPR001452; SH3.
CC DR Pfam: PF00612; IQ; 1.
CC DR Pfam: PF00663; myosin_head; 1.
CC DR Pfam: PF00018; SH3; 1.
CC DR PRINTS: PR00193; MYOSINHEAVY.
CC DR PRINTS: PR00452; SH3DOMAIN.
CC DR ProDom: PD000355; myosin_head; 1.
CC DR ProDom: PD000066; SH3; 1.
CC DR SMART: SM00242; MYSC; 1.
CC DR SMART: SK00242; MYSC; 1.
CC DR PROSITE: PS50096; IQ; 1.
CC DR PROSITE: PS50002; SH3; 1.
CC DR Myosin; ATP-binding; Actin-binding; Calmodulin-binding; SH3 domain;
CC Multigene family.
CC KW DOMAIN 1 679 MYOSIN HEAD-LIKE.
CC FT DOMAIN 695 724 IQ.
FT DOMAIN 1052 1109 SH3.
FT DOMAIN 581 591 ACTIN-BINDING (POTENTIAL).

FT NP BIND 112 119 ATP (POTENTIAL).
SQ SEQUENCE 1109 AA; 127040 MW; 0C1C74BA3705905 CRC64;
Query Match 5.3%; Score 100; DB 1; Length 1109;
Best Local Similarity 22.5%; Pred. No. 10;
Matches 64; Conservative 41; Mismatches 91; Indels 88; Gaps 15;

QY 47 GADPAKQKSELEF-----KFKRKLSEPYTSKSGKTYGHRKHLFSDQYRELQRYA 102
DB 476 GADDTLLQKQOMQSGHEHN-----SWNGQFLHHYAGKVSYDMQGFCEKRR 523
QY 103 EEVGIPTASGMDMAVEPLHNLVPPFKGSGDTNPPLEKTKAKGRPMVSSGQSM 162
DB 524 D---VLF---MD--LIELMQSSLEPIK-----SLFPEMLQADKKGRPTTAGSKIK-- 566
QY 163 DTMKQVQVYKPLPMPNFCPLQTSAY-----PLOPEDVNLRLVISEYOKLPPDIP 211
DB 567 ---KQANDLVSTL-----MKCTPHYTRCKPKNETKPPROMEEERVAGHOVEYGLKENIR 617
QY 212 IGVSGETGIAISVAVALGAKVLEHRTTDK---TWKSDHSASLEPGELAEVRSVR 267
DB 618 VRKGY-----AYRIRFOKFLQRYAILTKATWPMQGEKQGVL-----HLLQSVN 663
QY 268 LVERALSPYTKQLLPCEMACNEKKGKSVAKVKIPCEGTLITMDM 311
DB 664 M-----DSDQF-----QGRSKVP-ITAPESFLLEEM 690

RESULT 15

GPR_BACME STANDARD; PRT; 371 AA.
ID GPR_BACME
AC P23321;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Germination protease precursor (BC 3.4.24.78) (Spore protease) (GPR
DE endopeptidase) (Germination proteinase).
GN GPR.
OS Bacillus megaterium.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_Taxid=1404;
RN (1)
RP SEQUENCE FROM N.A. AND SEQUENCE OF 16-31.
RX STRAIN=ATCC 12872 / OMDC B1551;
RX MEDLINE=91100296; PubMed=1840582;
RA Suseman M.D., Setlow P.;
RT "Cloning, nucleotide sequence, and regulation of the Bacillus
RT subtilis gpr gene, which codes for the protease that initiates
RT degradation of small, acid-soluble proteins during spore
RT germination";
RT J. Bacteriol. 173:291-300(1991).
RL (2)
RN CHARACTERIZATION.
RP MEDLINE=98422459; PubMed=9748439;
RA Nessi C., Jedrzejewski M.J., Setlow P.;
RT "Structure and mechanism of action of the protease that degrades
RT small, acid-soluble spore proteins during germination of spores of
RT Bacillus species";
RT J. Bacteriol. 180:5077-5084(1998).
RL (3)
RN X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS) OF ZYMOCEN P46.
RP MEDLINE=20323395; PubMed=1864493;
RA Ponuraj K., Rowland S., Nessi C., Setlow P., Jedrzejewski M.J.;
RT "Crystal structure of a novel germination protease from spores of
RT Bacillus megaterium: structural arrangement and zymogen activation";
RT J. Mol. Biol. 300:11-10(2000).
CC -1- FUNCTION: Initiates the rapid degradation of small, acid-soluble
CC proteins during spore germination.
CC -1- CATALYTIC ACTIVITY: Endopeptidase action with P4 Glu or Asp, P1
CC preferably Glu > Asp, P1' hydrophobic and P2' Ala.
CC -1- SUBUNIT: Homotetramer.
CC -1- DEVELOPMENTAL STAGE: GPR transcription occurs during sporulation
CC in forespore first by sigma-F and then by sigma-G.

Query Match 5.2%; Score 98.5; DB 1; Length 371;
Best Local Similarity 20.9%; Pred. No. 3.4;
Matches 63; Conservative 44; Mismatches 120; Indels 75; Gaps 12;

```
Search completed: May 6, 2004, 09:05:03
Job time : 12.4405 secs
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 08:59:19 ; Search time 35.1395 Seconds

(Without alignments)
3223.466 Million cell updates/sec

Title: US-09-930-440B-6
Perfect score: 1887
Sequence: 1 MPELEBLCPGRWVGQHPF.....EEDTIMEELVDNHGKIKS 359

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : SPTRMBL_25:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriaph:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1810	95.9	359	11	Q99J77 mus musculi
2	1804	95.6	359	11	Q99J77 mus musculi
3	854	45.3	372	5	Q9V674 drosophila
4	547.5	29.0	280	16	Q8K655 chlorodum
5	534.5	28.3	341	2	Q87186 streptococ
6	534.5	28.3	341	16	Q8E501 streptococ
7	532.5	28.2	339	2	Q9ALW6 galact streptococ
8	532.5	28.2	339	2	Q93T72 galact streptococ
9	532.5	28.2	341	2	Q9RPO0 galact streptococ
10	532.5	28.2	341	16	Q8DZC3 galact streptococ
11	526	27.9	749	16	Q82UC3 galact streptococ
12	524	27.8	311	16	Q82HY4 galact streptococ
13	518	27.5	334	2	Q7X523 campylobact
14	516.5	27.4	333	16	Q8DD27 vibrio vuln
15	514	27.2	361	2	Q8KMA2 oskna2 pseudomonas
16	512	27.1	312	16	Q9AK45 oskna45 streptomyces

17	510.5	27.1	346	2	Q46675
18	510	27.0	338	2	Q9RDX5
19	509.5	27.0	351	16	Q893U6
20	507.5	26.9	753	16	Q8E504
21	506.5	26.8	341	2	Q9A016
22	500.5	26.5	641	16	Q89HL9
23	497	26.3	754	16	Q7U964
24	487	25.8	334	16	Q9PMX2
25	472	25.0	333	16	Q7U911
26	472	25.0	346	2	Q93N02
27	461.5	24.5	355	16	Q89HL8
28	437.5	23.2	286	17	Q8TUL5
29	435.5	23.1	344	16	Q8F508
30	435	22.9	346	16	Q8A711
31	431.5	22.9	357	16	Q87T70
32	419	22.2	350	16	Q97H26
33	410	21.7	351	2	Q8KH52
34	393.5	20.9	352	2	Q9R9S2
35	390.5	20.7	351	16	Q8R1B8
36	384.5	20.4	351	2	Q8KH54
37	382.5	20.3	349	16	Q87265
38	381.5	20.2	332	16	Q8F330
39	374.5	19.8	354	16	Q7V953
40	367.5	19.5	343	2	Q93J78
41	363.5	19.3	354	2	Q939M1
42	360.5	19.1	229	2	Q9F9F4
43	358	19.0	356	16	Q9A4G6
44	356.5	18.9	343	16	Q9PMY2
45	356	18.9	343	2	Q9EU02

ALIGNMENTS

RESULT 1
Q99J77 PRELIMINARY; PRT; 359 AA.
ID Q99J77;
AC Q99J77;
DT 01-JUN-2001 (T-EMBLrel. 17, Created)
DT 01-JUN-2001 (T-EMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Similar to N-acetylneuraminic acid phosphate synthase, sialic acid synthase.
DE
DE
GN NANS OR 463241804RIK OR SAS.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Skin;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RA "Analysis of the mouse transcriptome based on functional annotation of
RA Nature 420:563-573 (2002)."
RT 60,770 full-length cDNAs."
RL Nature 420:563-573 (2002).
EMBL: BC003507; AA003507.1; -
EMBL: AK076290; BA036290.1; -
HSSP: P19614; JUIA.
DR MGD; MGI:2149820; Nans.
DR GO; GO:0005829; C:cytosol; IDA.
DR GO; GO:0019007; F:N-acetylneuraminic acid phosphate synthase . . . ; IDA.
DR InterPro; IPR006013; AntifreezeZell.
DR InterPro; IPR006014; AntifreezeZell.
DR InterPro; IPR006190; AntifreezeZell.
DR InterPro; IPR006144; Neuh.
DR Pfam; PF01354; AntifreezeZell.
DR Pfam; PF03102; Neuh; 1.

DR PRINTS; PRO0357; ANTIFREEZII.
 DR PRODOM; PD003258; AntifreezeII; 1.
 DR PROSITE; PS50844; AFP LIKE; 1.
 SQ SEQUENCE 359 AA; 40024 MW; 4C66CB883558A373 CRC64;

Query Match 95.9%; Score 1810; DB 11; Length 359;
 Best Local Similarity 94.4%; Pred. No. 8.2e-141;
 Matches 339; Conservative 11; Mismatches 9; Indels 0; Gaps 0;

QY 1 MPELELCGRWVGSGHPDPTIIEIGQNHQGDIDVAKMIRTAKECGADCAKFOKSELEF 60
 DB 1 MPELELCGRWVGSGHPDPTIIEIGQNHQGDIDVAKMIRTAKECGADCAKFOKSELEF 60
 QY 61 KFNKRALERPPTYSKSHGKTYGKHLEPESHQVRELQRYAEVGIPTSGMDENAVE 120
 DB 61 KFNKRALERPPTYSKSHGKTYGKHLEPESHQVRELQRYAEVGIPTSGMDENAVE 120
 QY 121 FLEHLNVPFVKVSGDPTNNPYLEKTAKKGRPMVSSGMQSDMTKQVQIVKPLNPNFC 180
 DB 121 FLEHLNVPFVKVSGDPTNNPYLEKTAKKGRPMVSSGMQSDMTKQVQIVKPLNPNFC 180
 QY 181 FLOCTSAVPLQPEDVNLKRVISEYQKLPDPIPIGSGHETGIALSYAAVALGAKVLEHIT 240
 DB 181 FLOCTSAVPLQPEDVNLKRVISEYQKLPDPIPIGSGHETGIALSYAAVALGAKVLEHIT 240
 QY 241 LDKTKGSDHSASLEPPELAEIVRSVRLVERALGSPPTKQLPCEMACNEKLGKSVAVK 300
 DB 241 LDKTKGSDHSASLEPPELAEIVRSVRLVERALGSPPTKQLPCEMACNEKLGKSVAVK 300
 QY 301 IPEGITLTFMDLTVKVGEPYAPPEPDIENLVGKVLVTVVEEDDTIMEELVDNKGKITS 359
 DB 301 IPEGITLTFMDLTVKVGEPYAPPEPDIENLVGKVLVTVVEEDDTIMEELVDNKGKITS 359

RESULT 2

Q9JUH0 PRELIMINARY; PRT; 359 AA.
 AC Q9JUH0; 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
 DE N-acetylneuraminic acid 9-phosphate synthetase.
 GN NANS OR SAS.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Sutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;

RP SEQUENCE FROM N.A.
 RX MEDLINE=20334323; PubMed=10873558;
 RA Nakata D., Close B.E., Colley K.J., Matsuda T., Kitajima K.;
 RT "Molecular cloning and expression of the mouse N-acetylneuraminic acid
 RT 9-phosphate synthase which has not the deammonneuraminic acid (KDN)
 RT 9-phosphate synthase activity";
 RL Biochem Biophys Res Commun. 273:642-648(2000).
 DR EMBL; AB041263; BA98131.1; -;
 DR PIR; JC7321; JC7321.
 DR HSSP; P19614; J21A.
 DR MGD; MGJ:2149820; Mans.
 DR GO; GO:0005829; C:cytosol; IDA.
 DR GO; GO:0019007; P:N-acetylneuraminic acid phosphate synthase . . ; IDA.
 DR InterPro; IPR006190; Antifreeze_Like.
 DR InterPro; IPR004144; Neus.
 DR Pfam; PF03102; Neus; 1.
 DR PROSITE; PS50844; AFP_LIKE; 1.
 SQ SEQUENCE 359 AA; 39994 MW; 5FEB7D40C558A373 CRC64;

Query Match 95.6%; Score 1804; DB 11; Length 359;
 Best Local Similarity 94.2%; Pred. No. 2.5e-140;
 Matches 338; Conservative 11; Mismatches 10; Indels 0; Gaps 0;

QY 1 MPELELCGRWVGSGHPDPTIIEIGQNHQGDIDVAKMIRTAKECGADCAKFOKSELEF 60

DB 1 MPELELCGRWVGSGHPDPTIIEIGQNHQGDIDVAKMIRTAKECGADCAKFOKSELEF 60
 QY 61 KFNKRALERPPTYSKSHGKTYGKHLEPESHQVRELQRYAEVGIPTSGMDENAVE 120
 DB 61 KFNKRALERPPTYSKSHGKTYGKHLEPESHQVRELQRYAEVGIPTSGMDENAVE 120
 QY 121 FLEHLNVPFVKVSGDPTNNPYLEKTAKKGRPMVSSGMQSDMTKQVQIVKPLNPNFC 180
 DB 121 FLEHLNVPFVKVSGDPTNNPYLEKTAKKGRPMVSSGMQSDMTKQVQIVKPLNPNFC 180
 QY 181 FLOCTSAVPLQPEDVNLKRVISEYQKLPDPIPIGSGHETGIALSYAAVALGAKVLEHIT 240
 DB 181 FLOCTSAVPLQPEDVNLKRVISEYQKLPDPIPIGSGHETGIALSYAAVALGAKVLEHIT 240
 QY 241 LDKTKGSDHSASLEPPELAEIVRSVRLVERALGSPPTKQLPCEMACNEKLGKSVAVK 300
 DB 241 LDKTKGSDHSASLEPPELAEIVRSVRLVERALGSPPTKQLPCEMACNEKLGKSVAVK 300
 QY 301 IPEGITLTFMDLTVKVGEPYAPPEPDIENLVGKVLVTVVEEDDTIMEELVDNKGKITS 359
 DB 301 IPEGITLTFMDLTVKVGEPYAPPEPDIENLVGKVLVTVVEEDDTIMEELVDNKGKITS 359

RESULT 3

Q9VG74 PRELIMINARY; PRT; 372 AA.
 ID Q9VG74; 095VY1;
 AC Q9VG74; 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
 DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
 DE CG5232 protein (NeusAc synthase).
 GN NEUSAC OR CG5232.
 OS Drosophila melanogaster (fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pelecyota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RP SEQUENCE FROM N.A.
 RX STRAIN=Berkeley;
 RX MEDLINE=20196006; PubMed=10711132;
 RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.B., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.E., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter B.G., Heit G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Piankocch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Botkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
 RA Butts K.C., Butman D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.B., Downes M., Dugan-Rocha S., Dunov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Fertire W.S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
 RA Jaitani M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimel L.W., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusken D.R., Paclob J.M.,
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shie B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svendsen C., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.N., Weissbach J.,

QY 249 DHSASLEPGELAAELVRSVR LVERALGSP TKQLLPCEMACNEKL 291
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB 235 DQAASVEISGMSRLVSNIRDI EKALGDGVKRVYDGEAARKKL 277

RESULT 5

ID	087186	PRELIMINARY;	PRT;	341 AA
----	--------	--------------	------	--------

01-NOV-1998 (TReMBLrel. 08. Created)
 01-NOV-1998 (TReMBLrel. 08. Last sequence update)
 01-JUN-2003 (TReMBLrel. 24. Last annotation update)
 NEUB protein.
 GN
 DE
 Streptococcus agalactiae.
 OC
 Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 Streptococcus.
 CC
 NCBI_Textid=1311;
 [1]
 SEQUENCE FROM N.A.
 RA Yamamoto S., Miyake K., Iijima S.;
 RT "Identification and Characterization of cps (capsular polysaccharide)
 RT Genes from Streptococcus agalactiae Type Ia,"
 RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
 RN
 RN
 [2]
 SEQUENCE FROM N.A.
 RA Miyake K., Yamamoto S., Koike Y., Watanabe M., Iijima S.;
 RT "Molecular Characterization of Type-Specific Capsular Polysaccharide
 RT Biosynthesis Genes of Streptococcus agalactiae Type Ia,"
 RL Submitted (JUN-1999) to the EMBL/Genbank/DBJ databases.
 RN
 RN
 EMBL: AB017355; BAA33753.1; -
 DR EMBL: AB028895; BAA2287.1; -
 DR GO: GO:0016051; P:carbohydrate biosynthesis; IEA.
 DR InterPro: IPR006014; Antifreeze_dom.
 DR InterPro: IPR006190; Antifreeze_dom.
 DR InterPro: IPR004144; Neub.
 DR Pfam: PF01354; Antifreeze; 1.
 DR Pfam: PF01302; Neub; 1.
 DR PROSITE: PS50644; APF_LIKE; 1.
 SEQUENCE 341 Aa; 38091 MW; 176208F82A2EA340 CAC664;

ID	PRELIMINARY;	341 AA.
Q8E501	PRT;	
Q8E502		

RESULT 7

ID	Q9ALW6	PRELIMINARY;	PRT;	339	AA
----	--------	--------------	------	-----	----

RESULT 6
Q8E501

DR Pfam: PF01354; Antifreeze; 1.
 DR Pfam: PF03102; Neut; 1.
 DR PROSITE: PS50844; AFP_LIKE; 1.
 SQ SEQUENCE 341 AA; 38033 MW; 148B756C2B2BF314 CRC64;

Query Match: 28.2%; Score 532.5; DB 2; Length 341;
 Best Local Similarity 36.8%; Pred. No. 1.6e-35;
 Matches 125; Conservative 66; Mismatches 136; Indels 13; Gaps 6;

QY 20 PIIAIGQNHQGDLDVAKRMIMARECGADCAKQKSELEFENKALERPYTSKSGWK 79
 DB 3 YIIAIGCNHNDIMLAKKVVAVASCGVDVAFQFTFAEKLSKFAKAEY-QKATTTGT 61
 QY 80 TYG--EKKRLBPSHDQRELQRYAEVGIFFIASGNDMAVEFIHELVNPFKVGSDT 137
 DB 62 ADSQLEMTKRLBLSFEYELMRDVAISKGVETSTPDEBSLEFLISTDMPYKIPSGEI 121
 QY 138 NNPFYLEKTKAKGRPMVTSISGMSMDTKQVQIVKPLNPN---FCFLQCTSAVPLQPE 193
 DB 122 TNLPILEKIGKQKQKVLSTGMA---VMEIHQAVNILLRNGTTDISILHCTTEYPTEYP 178
 QY 194 DVNLKRVISEYQKLPDIPIGYSGHETGIAISVAVALGAKYLERHTLDTKTKSGDSAS 253
 DB 179 SLNLNVHTLTKDEFKDLITIGYSDHSIGSEVPDIAAAMGAIEIKHFTLDTNMEGPDHKS 238
 QY 254 LEPGELAEIVRSVLRALGSPTRKOLLPCENACNEKLGKSVAVAKYIPEGTILTMMLT 313
 DB 239 ATPDILALVKGRIEQLGRFEKIPDVEEKNNIVAKSVVALKPKIKGDIYSIENIT 298
 QY 314 VKVGEF-KAYPEDIFFNLVGRKVLTVVEEDDTIMEELVDN 352
 DB 299 VK--RPGNGISPMNWYDILGQEAQDDFEDEVIRDSRFEN 336

RESULT 10

Q8DZE3 PRELIMINARY; PRT; 341 AA.
 AC Q8DZE3;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-JUN-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE N-acetylneuraminic acid synthetase Neut.
 GN NEU8 OR SAG1161.
 OS Streptococcus agalactiae (serotype V).
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OC NCBI_Taxid=216466;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=2603 V/R / Serotype V.
 RX MEDLINE=22222988; PubMed=12200547;
 RA Teitelin H., Masiyuan V., Cieslewicz M.U., Eisen J.A., Peterson S.,
 RA Messels M.R., Paulsen I.T., Nelson K.E., Margarit I., Read T.D.,
 RA Madoff L.C., Wolf A.M., Beaman M.J., Brinkac L.M., Daugherty S.C.,
 RA Debay R.T., Dackiw A.S., Kolonay J.F., Madupu R., Lewis M.R.,
 RA Padure D., Fedorova N.B., Scanlan D., Khouri H., Mulligan S.,
 RA Carv H.A., Cline R.T., Van Aken S.E., Gill J., Scarselli M., Mora M.,
 RA Jacobini E.T., Breton C., Galli G., Mariani W., Vegni F., Malone D.,
 RA Rinaldo D., Rappuoli R., Telford J.L., Kaeper D.L., Grandi G.,
 RA Fraser C.M.;
 RT "Complete genome sequence and comparative genomic analysis of an
 RT emerging human pathogen, serotype V Streptococcus agalactiae.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:12391-12396(2002).
 DR EMBL: AB014425; J04651.
 DR PIR: T4651; T4651.
 DR TIGR: SAG1161;
 DR GO: GO:0016051; P:carbohydrate biosynthesis, IEA.
 DR InterPro: IPR006014; Antifreeze_dom.
 DR InterPro: IPR006190; Antifreeze_like.
 DR InterPro: IPR004144; Neut.
 DR Pfam: PF01354; Antifreeze; 1.
 DR Pfam: PF03102; Neut; 1.
 DR PROSITE: PS50844; AFP_LIKE; 1.

KM Complete proteome.
 SQ SEQUENCE 341 AA; 38033 MW; 148B756C2B2BF314 CRC64;

Query Match: 28.2%; Score 532.5; DB 16; Length 341;
 Best Local Similarity 36.8%; Pred. No. 1.6e-35;
 Matches 125; Conservative 66; Mismatches 136; Indels 13; Gaps 6;

QY 20 PIIAIGQNHQGDLDVAKRMIMARECGADCAKQKSELEFENKALERPYTSKSGWK 79
 DB 3 YIIAIGCNHNDIMLAKKVVAVASCGVDVAFQFTFAEKLSKFAKAEY-QKATTTGT 61
 QY 80 TYG--EKKRLBPSHDQRELQRYAEVGIFFIASGNDMAVEFIHELVNPFKVGSDT 137
 DB 62 ADSQLEMTKRLBLSFEYELMRDVAISKGVETSTPDEBSLEFLISTDMPYKIPSGEI 121
 QY 138 NNPFYLEKTKAKGRPMVTSISGMSMDTKQVQIVKPLNPN---FCFLQCTSAVPLQPE 193
 DB 122 TNLPILEKIGKQKQKVLSTGMA---VMEIHQAVNILLRNGTTDISILHCTTEYPTEYP 178
 QY 194 DVNLKRVISEYQKLPDIPIGYSGHETGIAISVAVALGAKYLERHTLDTKTKSGDSAS 253
 DB 179 SLNLNVHTLTKDEFKDLITIGYSDHSIGSEVPDIAAAMGAIEIKHFTLDTNMEGPDHKS 238
 QY 254 LEPGELAEIVRSVLRALGSPTRKOLLPCENACNEKLGKSVAVAKYIPEGTILTMMLT 313
 DB 239 ATPDILALVKGRIEQLGRFEKIPDVEEKNNIVAKSVVALKPKIKGDIYSIENIT 298
 QY 314 VKVGEF-KAYPEDIFFNLVGRKVLTVVEEDDTIMEELVDN 352
 DB 299 VK--RPGNGISPMNWYDILGQEAQDDFEDEVIRDSRFEN 336

RESULT 11

Q82UC3 PRELIMINARY; PRT; 749 AA.
 AC Q82UC3;
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Type III antifreeze protein: CBS domain: Neut family (EC 4.1.3.-).
 GN NR1570.
 OS Nitrosomonas europaea.
 OC Bacteria; Proteobacteria; Betaproteobacteria; Nitrosomonadales;
 OC Nitrosomonadaceae; Nitrosomonas.
 OC NCBI_Taxid=915;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 19718 / IPO 14298;
 RX MEDLINE=22586410; PubMed=12700255;
 RA Chain P., Iamerdin J.E., Iamert F.W., Regala W., Lao V., Land M.,
 RA Hauser L., Hooper A.B., Klotz M.G., Norton J., Sayavedra-Soto L.A.,
 RA Arciero D.M., Holmes N.G., Whitaker M.M., Alp D.J.;
 RT "Complete genome sequence of the ammonia-oxidizing bacterium and
 RT obligate chemolithoautotroph Nitrosomonas europaea.";
 RL J. Bacteriol. 185:2759-2773(2003).
 DR EMBL: BX31861; CAB5481.1;
 DR GO: GO:0016829; F:lyase activity; IEA.
 DR GO: GO:0016051; P:carbohydrate biosynthesis; IEA.
 DR InterPro: IPR006190; Antifreeze_dom.
 DR InterPro: IPR006190; Antifreeze_like.
 DR InterPro: IPR006190; Antifreeze_dom.
 DR InterPro: IPR006190; Antifreeze_like.
 DR Pfam: PF01354; Antifreeze; 1.
 DR Pfam: PF00571; CBS; 1.
 DR Pfam: PF03102; Neut; 1.
 DR PROSITE: PS50844; AFP_LIKE; 1.
 DR lyase; Complete proteome.
 SQ SEQUENCE 749 AA; 83356 MW; 284B8128108149A9 CRC64;

Query Match: 27.9%; Score 526; DB 16; Length 749;

Best Local Similarity 35.6%; Pred. No. 1.7e-34;
 Matches 115; Conservative 63; Mismatches 136; Indels 6; Gaps 5;

QY 13 VGGHCFIIAETIGONHGDLDVAKEMIMAKECGADCAKFOKSELEFFKRNKALERP72
 DB 125 IADQNPATFIIAEGVNNHNGDIIGLAKELVLAENAGADVCFQNRDLSLSYNOG--RAME 182
 QY 73 SKHSMWKG-TEBHHKHLFESHDOYRELQRYAEVGIFFTASGMDENAVEFLHNLNVPF 131
 DB 183 AGVDLGSQYTLDLINKFOLNHDELCOVFCYCRQODILPCTPMDVSAVHLDEYGLBAEK 242
 QY 132 VGGSDTNFPIYLEKTKAGKRPMTISSGMSQMDTMKQVQYQVRLNPNFCLQCTSAVPLQ 191
 DB 243 VASADNTNEMETLTKGTPPLICSTGMSSEADIKGSVDLRLGAPFALLHGNSTYPAE 302
 QY 192 PEVUNIRVISEYOKLFPDPIPIGSGHETGIAISVAVALGAKYLESHITLDTKTKWSDS 251
 DB 303 FQVNNLNTPLHMKQGLSTV-VGYSGRHGRSPVLAVALGARIVERKFTVDRSMGNDHX 361
 QY 252 ASLSEPGELAEVRSVRLVERALG-SPTKQLPCEMACNEGLGSSVYAKKIPBGTILTM 310
 DB 362 VSLPPEFAEMVQIRNIEBALQGGGRSLTQSEMINTLAKSLVINCDLSQGLIRDS 421
 QY 311 MLTVKGEPRKYPPEDEITNLVKG 333
 DB 422 MITVK-SFGQGLQPNRIDELAGK 443

RESULT 12

082HY4 PRELIMINARY; PRT; 311 AA.
 AC 082HY4
 DT 01-JUN-2003 (TEMBLrel. 24, Created)
 DT 01-JUN-2003 (TEMBLrel. 24, Last sequence update)
 DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)
 DE Putative N-acetylneuraminic acid (Neu5Ac) synthase.
 GN SAV3373.
 OS Streptomyces avermitilis.
 OC Bacteria; Actinobacteria; Actinobacteriales; Actinomycetales;
 OC Streptomycetaceae; Streptomycetaceae; Streptomyces.
 OC NCBI_TaxID=33903;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
 RX MEDLINE=21477403; PubMed=11572948;
 RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
 RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osone T.,
 RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
 RT "Genome sequence of an industrial microorganism Streptomyces
 RT avermitilis: deducing the ability of producing secondary
 RT metabolites";
 RT Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
 RL [2]
 RN (2)
 RP SEQUENCE FROM N.A.
 RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
 RX MEDLINE=22608306; PubMed=12692562;
 RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,
 RA Sakaki Y., Hattori M., Omura S.;
 RT "Complete genome sequence and comparative analysis of the industrial
 RT microorganism Streptomyces avermitilis";
 RT Nat. Biotechnol. 21:526-531(2003).
 RL [4]
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=VCI167;
 RX GO:GO:0004197; F:cysteine-type endopeptidase activity; IEA.
 DR GO:GO:0016051; P:carbohydrate biosynthesis; IEA.
 DR GO:GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro: IPR004144; Neut.
 DR InterPro: IPR000169; SHprot_acsite.
 DR Pfam: PF03102; NeutB; 1.
 DR PROSITE: PS00639; THOL_PROTEASE_HIS; 1.
 DR Complete proteome.
 SQ SEQUENCE 311 AA; 34162 MW; ELBOAFB7CAER400 CRC64;

Query Match 27.8%; Score 524; DB 16; Length 311;
 Best Local Similarity 38.9%; Pred. No. 7e-35;
 Matches 122; Conservative 53; Mismatches 115; Indels 24; Gaps 7;

QY 14 GGHCFIIAETIGONHGDLDVAKEMIMAKECGADCAKFOKSELEFFKRNKALERP72
 DB 15 GGHVYVYVGLGINHNELENAFLIDAAAGADAVKFCRTPEICTPRQMDIER-- 72
 QY 72 TSXSMWKG-TEBHHKHLFESHDOYRELQRYAEVGIFFTASGMDENAVEFLHNLNVPF 130
 DB 73 --DTMGRTVYIDYHRRVPEDEYRQIDEVAKSNITOMFASPMTEAVALKEKTDIPAH 130
 QY 131 KVGSDTNFPIYLEKTKAGKRPMTISSGMSQMDTMKQVQYQVRLNPNFCLQCTSAVPLQ 189
 DB 131 KVASLNDDELRLALRGRTVILSTG--ASTPKQLRHAVEVIGSDVILLCHASTVP 187
 QY 190 LQPEDNIRVISEYOKLFPDPIPIGSGHETGIAISVAVALGAKYLESHITLDTKTKWSD 249
 DB 188 AKAEEINIRVINTLTAEPVNPVPIGSGHETGIAISVAVALGATVEERITLDRMNGSD 247
 QY 250 HSASLSEPGELAEVRSVRLVERALGSPTRQLPCEMACNEKLGK--SVYAKKIPBGTIL 307
 DB 248 QASVPEPQSLTVLVDIRITKSLGDTGKYSBGLGPKKRLRRVGVAAEAI----- 301
 QY 308 TMDMLTVKGEPRK 321
 DB 302 -----AAAGEPVA 310

RESULT 13

07X523 PRELIMINARY; PRT; 334 AA.
 AC 07X523
 DT 01-OCT-2003 (TEMBLrel. 25, Created)
 DT 01-OCT-2003 (TEMBLrel. 25, Last sequence update)
 DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)
 DE Pmc.
 GN Pmc.
 OS Campylobacter coli.
 OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
 OC Campylobacteraceae; Campylobacter.
 OC NCBI_TaxID=195;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=VCI167;
 RX MEDLINE=964233180; PubMed=8825781;
 RA Query P., Doig P., Alm R.A., Burr D.H., Kinneela N., Trust T.J.;
 RT "Identification and characterization of genes required for post-
 RT translational modification of Campylobacter coli VCI167 flagellin";
 RL Mol. Microbiol. 19:365-378(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=VCI167;
 RA Logan S.M., Kelly J.F., Thibault P., Ewing C.P., Query P.;
 RT "Structural heterogeneity of carbohydrate modifications affects
 RT serospecificity of Campylobacter flagellins";
 RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=VCI167;
 RA Query P.M., Doig P., Alm R.A., Burr D.H., Kinneela N., Trust T.J.;
 RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=VCI167;
 RA Query P., Ewing C.P., Moran A.P., Trust T.J.;
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=VCI167;
 RA Logan S.M., Kelly J.F., Thibault P., Ewing C.P., Query P.;
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AY102521; AAM76277.1;
 SQ SEQUENCE 334 AA; 36905 MW; C415F8473D91DI56 CRC64;

Query Match 27.8%; Score 518; DB 2; Length 334;
 Best Local Similarity 36.6%; Pred. No. 2.4e-34;

Matches 123; Conservative 62; Mismatches 135; Indels 16; Gaps 5;

QY 21 IIAETGQNHQGLDVAKMTIMAKKCGADCAKQSELEFKKRLKALERTPTSK-HSMGX 79
 DB 6 IIAAGVNNHGDNLAKKLEIWAASGADPFAKQSFALCVSKAKKAAVQLKTTADE 65
 QY 80 TYGKRLHLEFSDQYREIQRVAEEVGIFFTASGNDMAVEFLHNLNVPFFKVGSGDTNN 139
 DB 66 SOLKMKLELDPMNAHQLLISHAKOGGIAFLSTADLESIEKLDLUGLEVFIIPIGSEITN 125
 QY 140 FPLYLKTAKKGRPMVIVSSGMSMDTKQYQYIV--KPLNPFCHLOCTSAIPLQPEBN 196
 DB 126 LFLYKIAKAKNKKIIISTGMSNIGIEALEVCEKGTORAKITLHCTTEYPADFNEVN 185
 QY 197 LRAVISEYOKLPDIPFGYSGHETGIAISVAVALGAKVLERHITLDTKWKSGDSHASLEP 256
 DB 186 LKAMOTLKNAF-NLDVGYSDHRTKIHISLTAALGASVLEKFTLIDKMSGPDHKSLEP 244
 QY 257 GBIAELVRSVRLVERALGSPYKQLLPCEKACNEKLGKSVAAKVIPEGITLTMMLTVKV 316
 DB 245 DEIQELCTKIREIESALGDGIXQASKSEKXNIEIAKSLVAKKIKKEIFSEENLITK- 303
 QY 317 GKKAPPPEDIPNL-----VGKVLVTEBEDTIME 347
 DB 304 -----RPASGISAMRYDEYLGKAKASXDYEDELHIE 334

RESULT 14
 Q8DD27 PRELIMINARY; PRT: 333 AA.

AC Q8DD27; Q8DD27, 23, Created)
 DT 01-OCT-2003 (TREMBLrel. 23, last sequence update)
 DT 01-JUN-2003 (TREMBLrel. 24, last annotation update)
 DE Sialic acid synthase.
 GN VVI0808.
 OS *Vibrio vulnificus*.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 OC Vibrionaceae; Vibrrio.
 OX NCBI_TaxID=672;
 RN SEQUENCE FROM N.A.
 RC STRAIN-CMCPS;
 RA Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,
 RA Choy H.E.;
 RT "Complete genome sequence of *Vibrio vulnificus* CMCPS."
 RL Submitted (DEC-2002) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AB016799; AAC09312.1; -
 DR GO: GO:0016051; P:carbohydrate biosynthesis; IEA.
 DR InterPro: IPR006014; Antifreeze dom.
 DR InterPro: IPR006190; Antifreeze_like.
 DR InterPro: IPR004144; Neut.
 DR Pfam: PF01354; Antifreeze; 1.
 DR Pfam: PF03102; Neut; 1.
 DR PROSITE: PS50844; AFP_LIKE; 1.
 DR Complete proteome.
 KM SEQUENCE 333 AA; 36504 MW; 419D1FIA07EBB2A4 CRC64;

Query Match 27.4%; Score 516.5; DB 16; Length 333;
 Best Local Similarity 39.0%; Pred. No. 3.2e-34;
 Matches 117; Conservative 53; Mismatches 125; Indels 5; Gaps 3;

QY 20 FIIAETGQNHQGLDVAKMTIMAKKCGADCAKQSELEFKKRLKALERTPTSKSM-G-78
 DB 6 FIIAAGVNNHGDNLAKKLEIWAASGADPFAKQSFALCVSKAKKAAVQLKTTADE 65
 QY 79 KTYGKRLHLEFSDQYREIQRVAEEVGIFFTASGNDMAVEFLHNLNVPFFKVGSGDTNN 138
 DB 66 ETOFEMKRLIELSYDPTLKSVCDDKGIPTFMSTPPEOSATPLDGLQA-VFKISGELT 124
 QY 139 NPFYLEKTAKKGRPMVIVSSGMSMDTKQYQYIVKPLNPF--FCFLQCTSAIPLQPEBN 195
 DB 125 NTFPLRLIASPAKRVILSTGMSNIGIEALEVCEKGTORAKITLHCTTEYPADFNEVN 184

QY 196 NLRAVISEYOKLPDIPFGYSGHETGIAISVAVALGAKVLERHITLDTKWKSGDSHASLE 255
 DB 185 NLKAMKTEQAFPDIPFGYSDHTLGTLPAAVALGAKVLEKFTLIDKMSGPDHKSASLE 244
 QY 256 FGBIAELVRSVRLVERALGSPYKQLLPCEKACNEKLGKSVAAKVIPEGITLTMMLTVK 315
 DB 245 FGBIAELVRSVRLVERALGSPYKQLLPCEKACNEKLGKSVAAKVIPEGITLTMMLTVK 304

RESULT 15
 Q8KNA2 PRELIMINARY; PRT: 361 AA.

AC Q8KNA2; Q8KNA2, 22, Created)
 DT 01-OCT-2002 (TREMBLrel. 22, last sequence update)
 DT 01-OCT-2002 (TREMBLrel. 22, last sequence update)
 DT 01-JUN-2003 (TREMBLrel. 24, last annotation update)
 DE Similar to Neut family.
 OS *Pseudomonas aeruginosa*.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Pseudomonadaceae; Pseudomonas.
 OX NCBI_TaxID=287;
 RN SEQUENCE FROM N.A.
 RP MEDLINE=22053227; PubMed=12057956;
 RX Raymond C.K., Sims E.H., Kas A., Spencer D.H., Kutyavlin T.V.,
 RA Ivey R.G., Zhou Y., Kaul R., Ciendomingo J.B., Olson M.V.;
 RT "Genetic Variation at the O-Antigen Biosynthetic Locus in *Pseudomonas*
 RT *aeruginosa*."
 RT J. Bacteriol. 184:3614-3622(2002).
 DR EMBL: AF498403; AA027589.1; -
 DR GO: GO:0016051; P:carbohydrate biosynthesis; IEA.
 DR InterPro: IPR006014; Antifreeze dom.
 DR InterPro: IPR006190; Antifreeze_like.
 DR InterPro: IPR004144; Neut.
 DR Pfam: PF01354; Antifreeze; 1.
 DR Pfam: PF03102; Neut; 1.
 DR PROSITE: PS50844; AFP_LIKE; 1.
 DR SEQUENCE 361 AA; 39425 MW; E63D3B9B5D076D CRC64;

Query Match 27.2%; Score 514; DB 2; Length 361;
 Best Local Similarity 35.8%; Pred. No. 5.7e-34;
 Matches 122; Conservative 55; Mismatches 132; Indels 32; Gaps 5;

QY 20 FIIAETGQNHQGLDVAKMTIMAKKCGADCAKQSELEFKKRLKALERTPTSKSMGX 79
 DB 10 FIIAAGVNNHGDNLAKKLEIWAASGADPFAKQSFALCVSKAKKAAVQLKTTADE 68
 QY 80 TYGKRLHLEFSDQYREIQRVAEEVGIFFTASGNDMAVEFLHNLNVPFFKVGSGDT 137
 DB 69 TESQMLMKKLELPKEMHFELOAHNHGIEPFIATFSDSLAFLAEKQLEFFKVPSSGL 128
 QY 138 NPFYLEKTAKKGRPMVIVSSGMSMDTKQYQYIV----- 172
 DB 129 TNGPLLMFAFKTGKRLIISTGMAITSEVQGLAIYAHLSGDNBERKMDDEVRLMSNSV 188
 QY 173 -KPLNPFCHLOCTSAIPLQPEBNLRAVISEYOKLPDIPFGYSGHETGIAISVAVALG 231
 DB 189 RMQLOGHVSLHCTSQYPTPPDEVNLLMDTLRF--GLAVGYSDHTGGLVPLAAVARG 246
 QY 232 AKVLERHITLDTKWKSGDSHASLEPGBIAELVRSVRLVERALGSPYKQLLPCEKACNEK 291
 DB 247 ACITIKHTTLDRSMGPHKASLEPGBIAELVRSVRLVERALGSPYKQLLPCEKACNEK 306
 QY 292 GKSVAAYKVIPEGITLTMMLTVKVEBPRAVPPEDIPNLVG 332
 DB 307 KQVVAARDIEAGMITRDILT-ARSGHGLPPTSLMEIAG 346

Search completed: May 6, 2004, 09:07:35
 Job time: 37.3895 secs

Thu May 13 11:53:08 2004

us-09-930-440b-6.ra1

Page 1

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: May 6, 2004, 08:59:54 ; Search time 14.9474 Seconds
(without alignments)
1239.929 Million cell updates/sec

Title: US-09-930-440B-6
Perfect score: 1887
Sequence: 1 MPELELCGRNVGSGHPCF.....EEDDTMEELVDNHGKIKS 359

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database: 1: /cgn2_6/ptodata/2/1aa/5A COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PTUS COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	1887	100.0	US-09-516-143A-4	Sequence 4, Appl1
2	336.5	17.8	US-09-455-406-21	Sequence 21, Appl1
3	113.5	6.0	US-09-107-532A-6711	Sequence 6711, Ap
4	100	5.3	US-09-648-004-10	Sequence 10, Appl1
5	93.5	5.0	US-09-134-000C-5846	Sequence 5846, Ap
6	93	4.9	US-09-252-991A-24216	Sequence 24216, A
7	93	4.9	US-08-907-166-2	Sequence 2, Appl1
8	93	4.9	US-09-391-340-2	Sequence 2, Appl1
9	92.5	4.9	US-09-343-681A-5022	Sequence 5022, Ap
10	92	4.9	US-09-498-520A-12	Sequence 12, Appl1
11	91.5	4.8	US-09-252-991A-25083	Sequence 25083, A
12	91	4.8	US-09-134-001C-3575	Sequence 3575, Ap
13	90	4.8	US-09-134-001C-4820	Sequence 4820, Ap
14	89	4.7	US-09-269-861A-8	Sequence 8, Appl1
15	88.5	4.7	US-09-134-000C-5157	Sequence 5157, Ap
16	88.5	4.7	US-09-337-913-1	Sequence 1, Appl1
17	88.5	4.7	US-08-750-524-1	Sequence 1, Appl1
18	88	4.7	US-08-492-027A-8	Sequence 15, Appl1
19	87	4.6	US-08-913-942-15	Sequence 26, Appl1
20	87	4.6	US-09-268-347-26	Sequence 6427, Ap
21	87	4.6	US-09-134-000C-6427	Sequence 16, Appl1
22	86.5	4.6	US-08-347-801-16	Sequence 142, Appl1
23	86.5	4.6	US-08-311-731A-142	Sequence 11591, A
24	86	4.6	US-09-489-039A-11591	Sequence 6, Appl1
25	85.5	4.5	US-07-768-286B-6	Sequence 3, Appl1
26	85.5	4.5	US-08-487-823B-3	Sequence 3, Appl1
27	85.5	4.5	US-08-997-040-3	Sequence 3, Appl1

28	85.5	4.5	382	2	US-09-203-237-3	Sequence 3, Appl1
29	85.5	4.5	760	3	US-08-928-941D-29	Sequence 29, Appl1
30	85.5	4.5	760	4	US-09-280-590A-29	Sequence 29, Appl1
31	85.5	4.5	760	4	US-09-892-398-29	Sequence 4939, Ap
32	85	4.5	789	4	US-09-134-000C-4939	Sequence 5890, Ap
33	85	4.5	1201	4	US-09-328-352-5890	Sequence 18, Appl1
34	84.5	4.5	303	4	US-08-928-941D-18	Sequence 18, Appl1
35	84.5	4.5	303	4	US-09-280-590A-18	Sequence 18, Appl1
36	84.5	4.5	303	4	US-09-892-398-18	Sequence 3, Appl1
37	84.5	4.5	323	2	US-09-019-216-3	Sequence 3, Appl1
38	84.5	4.5	392	1	US-09-249-241-3	Sequence 4, Appl1
39	84.5	4.5	392	1	US-07-768-286B-4	Sequence 7, Appl1
40	84.5	4.5	405	1	US-08-121-714-7	Sequence 7, Appl1
41	84.5	4.5	405	1	US-08-477-108A-7	Sequence 7, Appl1
42	84.5	4.5	405	2	US-08-477-112-7	Sequence 7, Appl1
43	84.5	4.5	405	5	PCT-US93-08322-7	Sequence 7, Appl1
44	84.5	4.5	415	1	US-07-911-531-19	Sequence 19, Appl1
45	84.5	4.5	415	1	US-07-693-636A-19	Sequence 19, Appl1

ALIGNMENTS

RESULT 1
US-09-516-143A-4
Sequence 4, Application US/09516143A
Patent No. 633182
GENERAL INFORMATION:
APPLICANT: Human Genome Sciences, Inc.
TITLE OF INVENTION: Human Glycosylation Enzymes
FILE REFERENCE: PF505PCT
CURRENT APPLICATION NUMBER: US/09/516,143A
PRIOR FILING DATE: 2000-03-01
PRIORITY APPLICATION NUMBER: 60/122,409
NUMBER OF SEQ ID NOS: 6
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 4
LENGTH: 359
TYPE: PRT
ORGANISM: Homo sapiens
US-09-516-143A-4

Query Match 100.0%; Score 1887; DB 4; Length 359;
Best Local Similarity 100.0%; Pred. No. 2.8e-193;
Matches 359; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MPELELCGRNVGSGHPCFIIAETGONHGGDLDVAKMIRAKKGGADCAKFKSELEF	60
DB	1	MPELELCGRNVGSGHPCFIIAETGONHGGDLDVAKMIRAKKGGADCAKFKSELEF	60
QY	61	KFNKKALERPPTYSKHSWGTGKHEKHLRFSHDYRELCRYAEVGIFFTAGGDMAYE	120
DB	61	KFNKKALERPPTYSKHSWGTGKHEKHLRFSHDYRELCRYAEVGIFFTAGGDMAYE	120
QY	121	FLHELVNPPFKVGGSTNNPPYLEKTAAGRPVVISGQSMQDMTKOVQIVELPNFC	180
DB	121	FLHELVNPPFKVGGSTNNPPYLEKTAAGRPVVISGQSMQDMTKOVQIVELPNFC	180
QY	181	FLQCTSAVPLQPEEDVLAIVISEYOKLPDIPIGYSGHETGIAISVAVALGANVLRHIT	240
DB	181	FLQCTSAVPLQPEEDVLAIVISEYOKLPDIPIGYSGHETGIAISVAVALGANVLRHIT	240
QY	241	LDTYKWSGHSASLEGEALAEVRSVLYERALGSTTKQLLPBMAENKLGKSVYAKK	300
DB	241	LDTYKWSGHSASLEGEALAEVRSVLYERALGSTTKQLLPBMAENKLGKSVYAKK	300
QY	301	IPGGTILTMQNTLVKGSBPRAVPPEDI FNLVGKQVAVTEEDDTIMEELVDNHGKIKS	359
DB	301	IPGGTILTMQNTLVKGSBPRAVPPEDI FNLVGKQVAVTEEDDTIMEELVDNHGKIKS	359

RESULT 2

Thu May 13 11:53:08 2004

us-09-930-440b-6.rai

Page 2

```

US-09-495-406-21
: Sequence 21. Application US/09495406
: Patent No.6503744
:
: GENERAL INFORMATION:
: APPLICANT: Gilbert, Michel
: APPLICANT: Wakarchuk, Warren W.
: APPLICANT: National Research Council of Canada
: TITLE OF INVENTION: Campylobacter glycosyltransferases for Biosynthesis of
: TITLE OF INVENTION: Campicosides and Ganglioside Mimics
: FILE REFERENCE: 019633-000110US
: CURRENT APPLICATION NUMBER: US/09/495,406
: CURRENT FILING DATE: 2000-01-31
: PRIOR APPLICATION NUMBER: US 60/118,213
: PRIOR FILING DATE: 1999-02-01
: NUMBER OF SEQ ID NOS: 35
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 21
:
: LENGTH: 346
: TYPE: PR1
: ORGANISM: Campylobacter jejuni
: FEATURE:
: OTHER INFORMATION: sialic acid synthase from C. jejuni OH4384 (CRF 8a
: OTHER INFORMATION: of LOS biosynthesis locus)
: US-09-495-406-21

```

Query Match	17.8%	Score 336.5	DB 4	Length 366
Best Local Similarity	29.1%	Pred. No. 2.4e-27		
Matches 36	Conservative 63	Mismatches 164	Indels 7	Gaps 4

```
QY      21 IIAELGONNGOLDVAKKERTIRMAKGGAGCAFOSESEFEKNRRLAL-RPFTSHSMMCK 79
        :::::::::::::::::::::
Db      19 VPEELGINNSSLBLAKIMVDAFSTGAKTIHQHIVADMSSAAKKITPENAKIS--- 75
        :::::::::::::::::::::
QY      80 TYGEHRKHLFESHDOYRELORYAEAVGIFFTASGDMDMAVEFLHEANVFKEYSGEPTTN 139
        :::::::::::::::::::::
Db      76 -IYEIMOKCALDYKDEIALKEYTEKIGLVLTSPPSBACANLEDMGSAFIIGSECNN 134
        :::::::::::::::::::::
QY      140 PYLELTKAKGRPMWISSGMOSENMTWKUYOJVKPLPNPFCELOCTSAVPLOPEDYNLRV 199
        :::::::::::::::::::::
Db      135 YELIKHILAFKPEMIVSTGMNSIBSKIPFKVLIDNBIEFFVALHTNLAPPHNYRLNA 194
        :::::::::::::::::::::
QY      200 ISEYOQLPDPIDIGSGHTGAIASVANALGAKULERITLTDTKWGSDBHASLEPGSL 259
        :::::::::::::::::::::
Db      195 MELKKEF-SCMWGISDHTTDLNACGAVALACVIERFTSMERSBGDIVCSMDTOAL 253
        :::::::::::::::::::::
QY      260 AEL-VRSAYLVERALGSPTKOLLPCMCACNEMLKGSVAKVMIPEGITILTMDLTVKGE 318
        :::::::::::::::::::::
Db      254 KELIIOSBOAMRGNNESKKAQKOQVITIDFAFASVGIKDKKGEEVLSMONIWIWRBG 313
        :::::::::::::::::::::
QY      319 PRAYRPEDIFNLVGKKUVALVEEDDTIMEE 348
        :::::::::::::::::::::
Db      314 LOGISABEPENITGKALKADIEDTOLSTE 343
```

RESULT 3
 US-09-107-532A-6711
 : Sequence 6711, Application US/09107532A
 Patent No. 6583275
 GENERAL INFORMATION:
 APPLICANT: LYNN A Doucelle-Stamm and David Bush
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
 ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
 NUMBER OF SEQUENCES: 7310
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: GENOMES THERAPEUTICS CORPORATION
 STREET: 100 Beaver Street
 CITY: Waltham
 STATE: Massachusetts
 COUNTRY: USA
 ZIP: 02354
 COMPUTER READABLE FORM:
 MEDIUM TYPE: CD/ROM ISO9660
 COMPUTER: PC

```

      OPERATING SYSTEM: <Unknown>
      SOFTWARE: ASCII
      CURRENT APPLICATION DATA:
        APPLICATION NUMBER: US/09/107,532A
        FILING DATE: 30-Jun-1998
      PRIOR APPLICATION DATA:
        APPLICATION NUMBER: 60/085,598
        FILING DATE: 14 May 1998
        APPLICATION NUMBER: 60/051571
        FILING DATE: July 2, 1997
      ATTORNEY/AGENT INFORMATION:
        NAME: Arditello, Pamela Deneke
        REGISTRATION NUMBER: 40,489
        REFERENCE/DOCKET NUMBER: GPC-012
      TELECOMMUNICATION INFORMATION:
        TELEPHONE: (781)893-5007
        TELEFAX: (781)893-8277
      INFORMATION FOR SEQ ID NO: 6711:
        SEQUENCE CHARACTERISTICS:
          LENGTH: 343 amino acids
          TYPE: amino acid
          TOPOLOGY: linear
          MOLECULE TYPE: protein
          HYPOTHETICAL: YES
          ORIGINAL SOURCE:
            ORGANISM: Enterococcus faecium
          FEATURE:
            NAME/KEY: misc feature
            LOCATION: (B) LOCATION 1...343
            SEQUENCE DESCRIPTION: SEQ ID NO: 6711:
US-09-107-532A-6711

```

ORGANISM: *Enterococcus faecium*
FEATURE:

LOCATION: (B) LOCATION 1...343

US-09-107-532A-6711

Query Match 6.0%; Score 113.5; DB 4; Length 343;
Best Local Similarity 22.4%; Pred No. 0.0017;
Matches 65; Conservative 48; Mismatches 102; Indels 75; Gaps 14;

Qy	10	GRWVGQHPHCITIAEIGNOGDDVAKREMTIRANECADCAKFOKSELEKFNKALER	63
Db	96	GSFVTAAPGCSI-----EG-IDQIRECARAKAGGA-----	KIRGAGFK 134
Qy	70	SYTSKHSWGKTYGKGRHLFEPSHQYRELQRYAEVGIFFTASGCDENAVEFLHIANVF	129
Db	135	PRISPYAR---QGLEEGGLKY-----IROADBEFGKIVTEWDBGHIMVAEYS--DI	183
Qy	130	EKVGSGDINNPPYLEKAKGRPMWVISGMSNDTMKO-----YQYIKPLNPNCFQO-	183
Db	184	LOIGARNMONKELISANGKISPKPGLKRGISG--TIENMLNAEYIINVEDSP-VIFER	240
Qy	184	-----CTSAYPLQPEDVNLTNIVISEYKLEPFDIGYSGHETG-----AISV	225
Db	241	GIRTYETATNTFPLSAVPLAKKLTTHPEVIAD-----PSHOTGIWEIETVPYAR	288
Qy	226	AAVALGA--KLEBHIITLTKWKSQDSHSLPEPGHIAELVRSVRLAL	273
Db	289	AGVAGSGADGMIVEIHPDPANAM--SDGQSLNEKYSIMNMEVALMKRAM	336

```

/ RESULT 4
/ US-09-648-004-10
/ Sequence 10, Application US/09648004
/ Patent No. 6498242
/ GENERAL INFORMATION:
/ APPLICANT: CHEN, QIONG
/ APPLICANT: THOMAS, STUART
/ APPLICANT: NAGARAJAN, VASANTHA
/ TITLE OF INVENTION: BIOLOGICAL METHOD FOR THE PRODUCTION OF ADIPIC ACID AND
/ TITLE OF INVENTION: INTERMEDIATES
/ FILE REFERENCE: CL-1341-A
/ CURRENT APPLICATION NUMBER: US/09/648,004
/ CURRENT FILING DATE: 2000-04-25
/ PRIOR APPLICATION NUMBER: 09/252,553
/ PRIOR FILING DATE: 1999-02-19
/ NUMBER OF SEQ ID NOS: 32

```

Thu May 13 11:53:08 2004

US-09-930-440b-6.rat

Page 3

SOFTWARE: Microsoft Office 97
SEQ ID NO 10
LENGTH: 547
TYPE: PRF
ORGANISM: Acinetobacter sp.
US-09-648-004-10

Query Match
Best Local Similarity 20.5%; Pred. No. 0.1;
Matches 82; Conservative 65; Mismatches 119; Indels 134; Gaps 24;

QY 4 ELET-----CGRWVGQHP-----CFI-IAEIQGNHGGDLDAKRMIR 41
DB 32 ELETQADKADKADVAGTWTNNKYPGALSDTEHLKYCSWDKELLOS-----LEIKKKYQ 87
QY 42 MAKECGADCAKFOKSELEFFKPKRKALERTYTSKSHNGKTYGKHRLLEFSHQYRE--L 98
DB 88 -----GPDVAKYIQVAA-----KHDIKSY-QFTAVQSAH-YNEADAL 125
QY 99 QYAAEVGIFETASGDEMAVEFLHELVNP-----FFKVGSGDGNPPYLEKTKKGRP 152
DB 126 WEVTTEYGDKTKRFL-ITNGLSLAPNLPNTKGINQFKEHLHTSRWP--DDVSTEGGR 182
QY 153 M-VISSGMSMDTMKVQYIVKELNPNFELQCTSAV-----PLQPEDVNLKRVISEQ 204
DB 183 VGVIGTG---STGVQVITAVAPLACHLTVFQBSAQYSPVIGNDPLSEBDVK-KIKUND 237
QY 205 KLEPDI---PIGSGHETGIAIVAAVALGAKYLEHITLDTKWK----- 246
DB 238 KMDGVNAGSALAGLINES---TVPAMSVAS--EKKAVERKAMQGGFRFNETFEDI 291
QY 247 GSDHSASLE---PEELAEIVSVRLVERALGSPKCOLPCMACNEKLKGS----- 294
DB 292 ATMEANITAEQNIRKIRAEIVADPAIAK-----LMPDLVAKRPLCSGYYNTFN 343
QY 295 -----VAAK-----VKIPGTLITMDMLTVKYG 317
DB 344 RDNVRLEDVKAHPIVEITENGVLKNGDFVELDMLICATG 383

RESULT 5

US-09-134-000C-5846
Sequence 5846, Application US/09134000C
Patent No. 6617156
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
FILE REFERENCE: 032796-032
CURRENT APPLICATION NUMBER: US/09/134,000C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/055,778
PRIOR FILING DATE: 1997-08-15
NUMBER OF SEQ ID NOS: 6812
SOFTWARE: PatentIn version 3.1
SEQ ID NO 5846
LENGTH: 389
TYPE: PRF
ORGANISM: Enterococcus faecalis
US-09-134-000C-5846

Query Match
Best Local Similarity 5.0%; Score 93.5; DB 4; Length 389;

Matches 73; Conservative 52; Mismatches 119; Indels 121; Gaps 17;

QY 38 RMIRAKGCGADCAKFOKSELE-----EFNKRKALERTYTSK-----HSM 77
DB 6 RMIOGKAWMIN-EQFQPAQVEVSEORIVAILFYNKAVDKGEGRIILPGFIIDHDM 64
QY 78 GKTYGKHRLLEFSHQYRELVQYAAEVGIFETASGDEMAVEFLHELVNPFFKVGSG--- 134
DB 65 ---HGADANJA--NHEPIKEMQAYLPEBGI---TAFLEPTTSTPFXOLHSEFVIGSFLIE 116

QY 135 --GOTTNPPYLEKTKKGRPMWISS-----GMSQMDTMKVQYIVKELNPNFCEL 182
DB 117 BDQGTNGAQLIGIRH--GPMISEFFSGSHNPBELLVKSVEFTFKKQELAK----- 165
QY 183 OCTSAVPLQPEDVNLKRVISEYOKLPPDIPIGVSGHETGIAISV-----AAVALG 231
DB 166 GHIKMLTAPENDVENALTYY-----CHEHDVVISIGHTAATYEQAMAAVEAG 213
QY 222 AKVLERHITLDTKTKGSHSASLEBGL-----AELVRSVRLVERALGSETKQL 281
DB 214 AK-----SFHTFTGEMDISHRKPTAVAAALDEETFAELIADGVHADYSL----- 259
QY 282 PCMACNEKLKGSVAAKVIPEGTLITMDMLTVKVGSEKAVP-PEDIFNLVGRKVLATVE 340
DB 260 -----VRVAKUKGQDYLAVTDSIAKGCQGVYPKPE-----KGIEMVID 301
QY 341 EDDTI 345
DB 302 EQNVV 306

RESULT 6

US-09-252-991A-24216
Sequence 24216, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196-116
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,768
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,180
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 24216
LENGTH: 375
TYPE: PRF
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-24216

Query Match
Best Local Similarity 4.9%; Score 93; DB 4; Length 375;

Matches 45; Conservative 32; Mismatches 67; Indels 56; Gaps 9;

QY 81 YGEHRLLEFSH-----DOYRELQYAAEVGIFETASGDEMAVEFLHELVNPFFKV 132
DB 162 YFDGHDHLDINEEYPRRFSHDYGATKYQAEGLVASADLGLVETALR-----FRFV 214
QY 133 GSGDTNPPYLEKTKKGRPMWISSGMSQMDTMKVQYIVKELNPNFCELQCTSAVPLQ 192
DB 215 GAGDTISIFPRMIOARKRRLILANGLRVD-----FTSVENLNDALFSLIAG--BF 265
QY 193 EDVNLKRVISEYOKL-FPDI-----PIGSGH-ETGIAISVAAV----- 228
DB 266 ALGKYTNISNGQPVFPWADVNVVWGQDLDPVG--GHLEPYANGYGLAALNEGVCILPGR 323
QY 229 -----ALGAKVLERHITLD 242
DB 324 PEPVLFRLGMAYMAKNFTLD 343

RESULT 7

US-08-907-166-2
Sequence 2, Application US/08907166
Patent No. 5948666
GENERAL INFORMATION:
APPLICANT: Callen, Walter
TITLE OF INVENTION: ISOLATION AND IDENTIFICATION OF NOVEL POLYMERASES
FILE REFERENCE: 09010/027001

Thu May 13 11:53:08 2004

us-09-930-440b-6.rai

Page 4

CURRENT APPLICATION NUMBER: US/08/907,166
CURRENT FILING DATE: 1997-08-06
NUMBER OF SEQ ID NOS: 12
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO 2
LENGTH: 868
TYPE: PRT
ORGANISM: Ammonifex degensis
US-08-907-166-2

Query Match 4.9%; Score 93; DB 2; Length 868;
Best Local Similarity 19.8%; Pred. No. 1.2;
Matches 72; Conservative 55; Mismatches 102; Indels 134; Gaps 17;

QY 51 AKFQSELEFFKFKRALKERPYTSKSHWGTGEGHKEHLEFSDQYRELQRYAEVIGFTT 110
DB 57 AAFDCKSTTF-----RHAALELY---KAHRPATPDELROQFNLIKGV----- 95
QY 111 ASGMDENAVEFLHEINVEFFKVGSGDTNNF--PYLEKTAKKGRPMVVISGMSMDTKQV 168
DB 96 -----LTALNVVVEKEGFEADDLIGTLVRAEKGMOCLIVTG--DLDAL--- 139
QY 169 YQIVKELNPFCLQCTSAVPLQPE-----DYNLRVISEYOKL-----PPDIPIGY 214
DB 140 -QIVSPITTVLMRKGISLAVFNEAEVKAERFGVTPRQLPDKLAGSDASDNIPGLP--- 195
QY 215 SGHETGIAISVAANVALGAKVLERHTITLDTKWSGDSASLPEGLAEIVRSVLRERALG 274
DB 196 -----GIGPKYA-----SRLLQSHQLEKLE---SKRFFPAKLEETLE--RHKERAVL 239
QY 225 SPTKQLLPCMACNENK-----GKSVAAKVI-----PEGTILITMDML 312
DB 240 AKKALIRRDVPLEEETIRPWPGPNTLATLEVFSLERLRTAKRLELFPFARLLSASGL 299
QY 313 T-----VKGEPK-----AYPP-----EDIFNLVVK 333
DB 300 TPSAVRVKVERPRELERLGELGROFPALAVPVLRRKATSSFLALCLGGEKVFLEEGP 359
QY 334 KVL 336
DB 360 EVL 362

RESULT 8

US-09-391-340-2
Sequence 2, Application US/09391340A
Patent No. 6492511
GENERAL INFORMATION:
APPLICANT: Callen, Walter
APPLICANT: Mather, Eric
TITLE OF INVENTION: ISOLATION AND IDENTIFICATION OF NOVEL POLYMERASES
FILE REFERENCE: 09010/027001
CURRENT APPLICATION NUMBER: US/09/391,340A
CURRENT FILING DATE: 1999-09-07
EARLIER APPLICATION NUMBER: US 08/907,166
NUMBER OF SEQ ID NOS: 12
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 2
LENGTH: 868
TYPE: PRT
ORGANISM: Ammonifex degensis
US-09-391-340-2

Query Match 4.9%; Score 93; DB 4; Length 868;
Best Local Similarity 19.8%; Pred. No. 1.2;
Matches 72; Conservative 55; Mismatches 102; Indels 134; Gaps 17;

QY 51 AKFQSELEFFKFKRALKERPYTSKSHWGTGEGHKEHLEFSDQYRELQRYAEVIGFTT 110
DB 57 AAFDCKSTTF-----RHAALELY---KAHRPATPDELROQFNLIKGV----- 95
QY 111 ASGMDENAVEFLHEINVEFFKVGSGDTNNF--PYLEKTAKKGRPMVVISGMSMDTKQV 168

DB 96 -----LTALNVVVEKEGFEADDLIGTLVRAEKGMOCLIVTG--DLDAL--- 139
QY 169 YQIVKELNPFCLQCTSAVPLQPE-----DYNLRVISEYOKL-----PPDIPIGY 214
DB 140 -QIVSPITTVLMRKGISLAVFNEAEVKAERFGVTPRQLPDKLAGSDASDNIPGLP--- 195
QY 215 SGHETGIAISVAANVALGAKVLERHTITLDTKWSGDSASLPEGLAEIVRSVLRERALG 274
DB 196 -----GIGPKYA-----SRLLQSHQLEKLE---SKRFFPAKLEETLE--RHKERAVL 239
QY 225 SPTKQLLPCMACNENK-----GKSVAAKVI-----PEGTILITMDML 312
DB 240 AKKALIRRDVPLEEETIRPWPGPNTLATLEVFSLERLRTAKRLELFPFARLLSASGL 299
QY 313 T-----VKGEPK-----AYPP-----EDIFNLVVK 333
DB 300 TPSAVRVKVERPRELERLGELGROFPALAVPVLRRKATSSFLALCLGGEKVFLEEGP 359
QY 334 KVL 336
DB 360 EVL 362

RESULT 9

US-09-543-681A-5022
Sequence 5022, Application US/09543681A
Patent No. 6605703
GENERAL INFORMATION:
APPLICANT: GARY BRETON
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILI
FILE REFERENCE: 2709,1002-001
CURRENT APPLICATION NUMBER: US/09/543,681A
CURRENT FILING DATE: 2000-04-05
PRIOR APPLICATION NUMBER: US 60/128,706
NUMBER OF SEQ ID NOS: 8344
SEQ ID NO 5022
LENGTH: 753
TYPE: PRT
ORGANISM: Proteus mirabilis
US-09-543-681A-5022

Query Match 4.9%; Score 92.5; DB 4; Length 753;
Best Local Similarity 20.8%; Pred. No. 1.1;
Matches 69; Conservative 58; Mismatches 105; Indels 99; Gaps 18;

QY 117 MAVEFLHEINVEFFKVGSGDTNNF-----LEKTAKKGRPMVVISGMS 161
DB 152 IAKDTLIDNFPVLYDQKETTLLPIYGEOSMEKKRTLELEKLABEKGAEVYQYLEN 211
QY 162 MDTKQ-----VYQIVKELNPFCLQCTSAVPLQPEQVNLRLVISEY----- 203
DB 212 KDSVKQSKTKKMRPIREKDPQWYTK-----YKLISQDINILFDSAYLVNPPATYILY 267
QY 204 -----GKLPF-----DIPIGYS---GEGTIAISVAANV---ALGA---KYLERH 238
DB 268 QIVKGDKNVPLQENSEFLNIYLLKKSADLHHQILNIIIEHNNSNTLSSYFNNLEKY 327
QY 239 I-TLDTKWSGDH-----SASLEP--GELAEIVRSVLRERALG--SPTKQLLPCMAC 287
DB 328 IDTLLKTPNSPQAMALANVYLKPNSSFNFEKALVKAAYVIGSPSKILLALYS 387
QY 288 N-EKLGKSVAAKVIPEGTILITMDLT-----VKV-----GEKXATPEDIPTLVNCKV 335
DB 388 NSEGVHONIRKAVSLKENI-TNDKLTGKSQRELVKTIYFDGASDYLLKKEIVNLRESV 446
QY 336 LV-----TVEED-----DTIMEELVDNH 353
DB 447 IKNSAGTINONISLAFYADLLIEEDVANN 477

[illegible]

	Query Match	4.7%;	Score 89;	DB 4;	Length 831;	
	Best Local Similarity	22.3%;	Pred. No. 3;			
	Matches	72;	Conservative	38;	Mismatches	105;
					Indels	108;
					Gaps	15;
Qy	54	QKSBLEFFKPKALERPPTSQHSNGKTYGEKRLHFEHQDRELRQRAAEVGIFFT---	110			
		:				
Db	216	RKDEKPKRLREKLAHEHMIAKISQALILREIPLEILELQKVEPVEBAKPLHLE	275			
		:				
Qy	111	-----ASGDEMVEFLH---ELNVPE---FKVSGGQTNNEPYLEKTA	148			
		:				
Db	276	FKSPLEKEIPKIKKEYOGKDLVQVEIETESQILVVSDFYVDDEKTKYSLDRNE	335			
		:				
Qy	149	-----KGRPMYISSGQMDTMKQVOYL--KPLN--PNEFLQCTSAVPLQ-----	192			
		:				
Db	336	IEEIPFRNKU-----TDDAKGIYHCHLEKGLTFPEVCFDARIYAAYLVNADQNPGLK	388			
		:				
Qy	193	-----EDVNLRY-----ISEYQKLPFDI-----PIGYSGHEGSI	221			
		:				
Db	389	GLYLKYDLPVEYEDVSLNIRGLFYLNKEMMRKAPFQEGQRLFELEILPLTPVLAQEHHTDI	448			
		:				
Qy	222	AISVAAYALGAKVLERHTLTDKTKGSDHSASLEPGE--LAELVRSVRLV---ERALGSP	277			
		:				
Db	449	QYDRBAL-----KENSLELGEQLEHILREIYVLAGEBFNLS-	486			
		:				
Qy	278	KQLLPCENACNEKLGKSVAKYK	300			
		:				
Db	487	RQL-----GVILEFKGLPVKKTK	506			
		:				

```

      RESULT 15
      US-09-134-000C-5157
      ; Sequence 5157, Application US/09134000C
      ; Patent No. 6617156
      ; GENERAL INFORMATION:
      ; APPLICANT: Lyph Doucette-Stamm et al
      ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
      ; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
      ; FILE REFERENCE: 032796-032
      ; CURRENT APPLICATION NUMBER: US/09/134.000C
      ; CURRENT FILING DATE: 1998-08-13
      ; PRIOR APPLICATION NUMBER: US 60/055,778
      ; PRIOR FILING DATE: 1997-08-15
      ; NUMBER OF SEQ ID NOS: 6812
      ; SOFTWARE: PatentIn version 3.1
      ; SEQ ID NO 5157
      ; LENGTH: 247
      ; TYPE: PRT
      ; ORGANISM: Enterococcus faecalis
      ;
      US-09-134-000C-5157

      Query Match          4.7%; Score 88.5; DB 4; Length 247;
      Best Local Similarity 22.0%; Pred. No. 0.48;
      Matches 49; Conservative 35; Mismatches 84; Indels 55; Gaps 11

      QY 34 DVAKR-----WIMAKGCGADCAKFKSKSEFEKFKKALERP-----YTSKHSWG 78
      :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
      DB 44 EVAKCFCSPSSIIIRAKKAKNTS---GYNELIVLAKKAHFCQPIPFETAPSFETINERC 99
      :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

      QY 79 KTYGEGKRLH--EFSHDYRELQRYAAEEVGIFFTASGQDEMAVEFLHLSINVEFFKVGSGD 136
      :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
      DB 100 RLAAHGKSHLPVILIGDFFSRHLAAVISEY---FNFGHGFSLITTAATHSIN-----SON 149

```


Thu May 13 11:53:08 2004

us-09-930-440b-6.rai

Page 7

```
QY 137 TNNF-----PYLEKTA---KGRPMVYISSGMOGMDYKQYQIVKPELNPFCF 181
Db 150 NQNFLLIILSHSGEKKYKLTALAKKXKPIISFVQAKN-STLGRDLADLVFSTDYSYSPF 208
QY 182 LQCTSAVPIQPEDVNLKRVISYQKLPFDIPIGYSGHFTGIAIS 224
Db 209 ---STVAQPCMFPGQTLITFELI---CAYLNHEDSIPIS 243
```

Search completed: May 6, 2004, 09:08:38
Job time : 16.9474 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:07:45 ; Search time 38.8108 Seconds
(Without alignments)
2567.492 Million cell updates/sec

Title: US-09-930-440b-6
Sequence: 1887
1 MPELELCGRWVGQHPGCF.....EEDDTIMEELVDNKGKINS 359

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1140673 seqs, 277566755 residues

Total number of hits satisfying chosen parameters: 1140673

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1887	100.0	359	US-09-984-205-4	Sequence 4, Appl1
2	1887	100.0	359	US-09-930-440b-6	Sequence 6, Appl1
3	1883	99.8	359	US-10-264-237-2597	Sequence 2597, Ap
4	544.5	28.9	338	US-09-767-041-24	Sequence 24, Appl1
5	524	27.8	311	US-10-156-761-10910	Sequence 10910, A
6	510.5	27.1	346	US-09-930-440b-8	Sequence 8, Appl1
7	381	20.2	123	US-10-106-698-6926	Sequence 6926, Ap
8	336.5	17.8	346	US-09-816-0284-35	Sequence 35, Appl1
9	336.5	17.8	346	US-10-303-161-35	Sequence 35, Appl1
10	336.5	17.8	346	US-10-303-118-35	Sequence 35, Appl1
11	336.5	17.8	346	US-10-303-128-35	Sequence 35, Appl1
12	336.5	17.8	346	US-10-303-134-35	Sequence 35, Appl1
13	336.5	17.8	346	US-10-303-162-35	Sequence 35, Appl1
14	303	16.1	340	US-09-882-227-490	Sequence 490, Appl
15	287	15.2	340	US-10-335-977-7626	Sequence 7626, Ap

16	283.5	15.0	314	US-10-335-977-7625	Sequence 7625, Ap
17	122.5	6.5	205	US-10-321-396C-4	Sequence 4, Appl1
18	119	6.3	352	US-10-369-493-18790	Sequence 18790, A
19	117.5	6.2	275	US-10-321-396C-2	Sequence 2, Appl1
20	116	6.1	352	US-10-369-493-20092	Sequence 20092, A
21	113.5	6.0	333	US-10-282-122A-57486	Sequence 57486, A
22	113	6.0	285	US-10-369-493-18119	Sequence 18119, A
23	110	5.8	135	US-10-321-396C-6	Sequence 6, Appl1
24	109	5.8	265	US-10-369-493-31568	Sequence 21568, A
25	109	5.8	337	US-10-369-493-9746	Sequence 9746, Ap
26	108.5	5.7	335	US-10-369-493-20181	Sequence 20181, A
27	108.5	5.7	333	US-10-369-493-7631	Sequence 2631, Ap
28	108.5	5.7	633	US-10-282-122A-63506	Sequence 63506, A
29	108	5.7	338	US-10-369-493-2930	Sequence 2930, Ap
30	107.5	5.7	337	US-10-282-122A-51516	Sequence 51516, A
31	107	5.7	276	US-10-369-493-22853	Sequence 22853, A
32	107	5.7	330	US-10-282-122A-67597	Sequence 67597, A
33	103.5	5.5	329	US-10-369-493-13704	Sequence 13704, A
34	101.5	5.4	335	US-10-369-493-8945	Sequence 8945, Ap
35	101.5	5.4	337	US-10-282-122A-53304	Sequence 53304, A
36	101.5	5.4	341	US-10-282-122A-57228	Sequence 57228, A
37	101	5.4	1175	US-10-369-493-6504	Sequence 6504, Ap
38	100	5.3	547	US-10-369-493-15488	Sequence 15488, A
39	100	5.3	547	US-10-369-493-15860	Sequence 15860, A
40	100	5.3	1108	US-10-087-192-1206	Sequence 1206, Ap
41	100	5.3	1189	US-10-282-122A-70920	Sequence 70920, A
42	97.5	5.2	640	US-10-369-493-16235	Sequence 16235, A
43	97.5	5.2	641	US-10-369-493-15488	Sequence 15488, A
44	97.5	5.2	641	US-10-369-493-15860	Sequence 15860, A
45	97.5	5.2	642	US-10-389-566-1089	Sequence 1089, Ap

ALIGNMENTS

RESULT 1
US-09-984-205-4
Sequence 4, Application US/09984205
Patent No. US20020137175A1
GENERAL INFORMATION:
APPLICANT: Coleman, Timothy A. et al.
TITLE OF INVENTION: Human Glycosylation Enzymes
FILE REFERENCE: PF505D1
CURRENT APPLICATION NUMBER: US/09/984,205
CURRENT FILING DATE: 2001-10-29
PCT/US00/05325
PRIOR APPLICATION NUMBER: 2000-03-01
PRIOR FILING DATE: 2000-03-01
PRIOR APPLICATION NUMBER: 09/516,143
PRIOR FILING DATE: 2000-03-01
PRIOR APPLICATION NUMBER: 60/122,409
PRIOR FILING DATE: 1999-03-02
NUMBER OF SEQ ID NOS: 6
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 4
LENGTH: 359
TYPE: PRT
ORGANISM: Homo sapiens
US-09-984-205-4

Query Match 100.0% Score 1887, DB 9; Length 359;
Best Local Similarity 100.0%; Pred. No. 1.5e-182;
Matches 359; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MPELELCGRWVGQHPGCFIIAIGQHQGLDVAKRMIRPAKCEGACAKFQKSELEF 60
DB 1 MPELELCGRWVGQHPGCFIIAIGQHQGLDVAKRMIRPAKCEGACAKFQKSELEF 60
QY KERRKALEPPTYSKSHKWTGHEHKKHLEPFSIDQREIOLRYAEVGIPTASGDMAYE 120
DB KERRKALEPPTYSKSHKWTGHEHKKHLEPFSIDQREIOLRYAEVGIPTASGDMAYE 120
QY FLHELVNPFKVGSDPTNNFPLEKTAKKGRPVYSSGQSHDTMKQVQIVKPLPNFC 180
DB FLHELVNPFKVGSDPTNNFPLEKTAKKGRPVYSSGQSHDTMKQVQIVKPLPNFC 180

```

Db      121  FLEHNVPPFKVSGDITNNPYLEKTAAGKGPWVIVSSGMSQMDTMKVYIVKELNENFC 180
Qy      181  FLOCTSAVPILOPEVDNIRVISEYOKLPDIPITIGSGHETGIAISVAANALCAKYLERHIT 240
Db      181  FLOCTSAVPILOPEVDNIRVISEYOKLPDIPITIGSGHETGIAISVAANALCAKYLERHIT 240
Qy      241  LDKTWKSGDHSASLEPGLAEIVRSVRLVERALGSPYKOLLPCEMACNEKLGKSVAVAKV 300
Db      241  LDKTWKSGDHSASLEPGLAEIVRSVRLVERALGSPYKOLLPCEMACNEKLGKSVAVAKV 300
Qy      301  IPEGITLTMDMITLVKVGSPKAYPPEDIFNLVKKVLTVEEDDTIMEELVNHGKJIKS 359
Db      301  IPEGITLTMDMITLVKVGSPKAYPPEDIFNLVKKVLTVEEDDTIMEELVNHGKJIKS 359

RESULT 2
US-09-930-440b-6
; Sequence 6, Application US/09930440B
; Patent No. US20020142386A1
; GENERAL INFORMATION:
; APPLICANT: Betenbaugh et al.
; TITLE OF INVENTION: Engineering Intracellular Signaling Pathways
; FILE REFERENCE: P05092P
; CURRENT APPLICATION NUMBER: US/09/930,440B
; PRIOR FILING DATE: 2001-08-16
; PRIOR APPLICATION NUMBER: 60/227,579
; PRIOR FILING DATE: 2000-08-25
; PRIOR APPLICATION NUMBER: 09/516,793
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/169,624
; PRIOR FILING DATE: 1999-12-08
; PRIOR APPLICATION NUMBER: 60/122,582
; PRIOR FILING DATE: 1999-03-02
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 6
; LENGTH: 359
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-930-440b-6

```

```

Query Match      100.0%; Score 1887; DB 9; Length 359;
Best Local Similarity 100.0%; Pred. No. 1.5e-182;
Matches 359; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      1  MPELELCPRGWVGQHPCEFIIEIGNHQDDLVAKRMIRMAKCGADCAKFOKSELEF 60
Db      1  MPELELCPRGWVGQHPCEFIIEIGNHQDDLVAKRMIRMAKCGADCAKFOKSELEF 60
Qy      61  KFNKALERPPTSGHSGKTYGHEKRLHESHDYRELORVAREVGIFFTASGMDMAVE 120
Db      61  KFNKALERPPTSGHSGKTYGHEKRLHESHDYRELORVAREVGIFFTASGMDMAVE 120
Qy      61  KFNKALERPPTSGHSGKTYGHEKRLHESHDYRELORVAREVGIFFTASGMDMAVE 120
Db      61  KFNKALERPPTSGHSGKTYGHEKRLHESHDYRELORVAREVGIFFTASGMDMAVE 120
Qy      121  FLEHNVPPFKVSGDITNNPYLEKTAAGKGPWVIVSSGMSQMDTMKVYIVKELNENFC 180
Db      121  FLEHNVPPFKVSGDITNNPYLEKTAAGKGPWVIVSSGMSQMDTMKVYIVKELNENFC 180
Qy      121  FLEHNVPPFKVSGDITNNPYLEKTAAGKGPWVIVSSGMSQMDTMKVYIVKELNENFC 180
Db      121  FLEHNVPPFKVSGDITNNPYLEKTAAGKGPWVIVSSGMSQMDTMKVYIVKELNENFC 180
Qy      181  FLOCTSAVPILOPEVDNIRVISEYOKLPDIPITIGSGHETGIAISVAANALCAKYLERHIT 240
Db      181  FLOCTSAVPILOPEVDNIRVISEYOKLPDIPITIGSGHETGIAISVAANALCAKYLERHIT 240
Qy      181  FLOCTSAVPILOPEVDNIRVISEYOKLPDIPITIGSGHETGIAISVAANALCAKYLERHIT 240
Db      181  FLOCTSAVPILOPEVDNIRVISEYOKLPDIPITIGSGHETGIAISVAANALCAKYLERHIT 240
Qy      241  LDKTWKSGDHSASLEPGLAEIVRSVRLVERALGSPYKOLLPCEMACNEKLGKSVAVAKV 300
Db      241  LDKTWKSGDHSASLEPGLAEIVRSVRLVERALGSPYKOLLPCEMACNEKLGKSVAVAKV 300
Qy      301  IPEGITLTMDMITLVKVGSPKAYPPEDIFNLVKKVLTVEEDDTIMEELVNHGKJIKS 359
Db      301  IPEGITLTMDMITLVKVGSPKAYPPEDIFNLVKKVLTVEEDDTIMEELVNHGKJIKS 359

```

```

RESULT 3
US-10-264-237-2597
; Sequence 2597, Application US/10264237

```

```

; Publication No. US20040009491A1
; GENERAL INFORMATION:
; APPLICANT: Birtse et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: P0131P1
; CURRENT APPLICATION NUMBER: US/10/264,237
; PRIOR FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/16450
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: US 60/205,515
; PRIOR FILING DATE: 2000-05-19
; NUMBER OF SEQ ID NOS: 2876
; SOFTWARE: Patent In Ver. 3.1
; SEQ ID NO 2597
; LENGTH: 359
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-264-237-2597

```

```

Query Match      99.8%; Score 1883; DB 15; Length 359;
Best Local Similarity 99.7%; Pred. No. 3.7e-182;
Matches 358; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy      1  MPELELCPRGWVGQHPCEFIIEIGNHQDDLVAKRMIRMAKCGADCAKFOKSELEF 60
Db      1  MPELELCPRGWVGQHPCEFIIEIGNHQDDLVAKRMIRMAKCGADCAKFOKSELEF 60
Qy      61  KFNKALERPPTSGHSGKTYGHEKRLHESHDYRELORVAREVGIFFTASGMDMAVE 120
Db      61  KFNKALERPPTSGHSGKTYGHEKRLHESHDYRELORVAREVGIFFTASGMDMAVE 120
Qy      61  KFNKALERPPTSGHSGKTYGHEKRLHESHDYRELORVAREVGIFFTASGMDMAVE 120
Db      61  KFNKALERPPTSGHSGKTYGHEKRLHESHDYRELORVAREVGIFFTASGMDMAVE 120
Qy      121  FLEHNVPPFKVSGDITNNPYLEKTAAGKGPWVIVSSGMSQMDTMKVYIVKELNENFC 180
Db      121  FLEHNVPPFKVSGDITNNPYLEKTAAGKGPWVIVSSGMSQMDTMKVYIVKELNENFC 180
Qy      121  FLEHNVPPFKVSGDITNNPYLEKTAAGKGPWVIVSSGMSQMDTMKVYIVKELNENFC 180
Db      121  FLEHNVPPFKVSGDITNNPYLEKTAAGKGPWVIVSSGMSQMDTMKVYIVKELNENFC 180
Qy      181  FLOCTSAVPILOPEVDNIRVISEYOKLPDIPITIGSGHETGIAISVAANALCAKYLERHIT 240
Db      181  FLOCTSAVPILOPEVDNIRVISEYOKLPDIPITIGSGHETGIAISVAANALCAKYLERHIT 240
Qy      181  FLOCTSAVPILOPEVDNIRVISEYOKLPDIPITIGSGHETGIAISVAANALCAKYLERHIT 240
Db      181  FLOCTSAVPILOPEVDNIRVISEYOKLPDIPITIGSGHETGIAISVAANALCAKYLERHIT 240
Qy      241  LDKTWKSGDHSASLEPGLAEIVRSVRLVERALGSPYKOLLPCEMACNEKLGKSVAVAKV 300
Db      241  LDKTWKSGDHSASLEPGLAEIVRSVRLVERALGSPYKOLLPCEMACNEKLGKSVAVAKV 300
Qy      301  IPEGITLTMDMITLVKVGSPKAYPPEDIFNLVKKVLTVEEDDTIMEELVNHGKJIKS 359
Db      301  IPEGITLTMDMITLVKVGSPKAYPPEDIFNLVKKVLTVEEDDTIMEELVNHGKJIKS 359

```

```

RESULT 4
US-09-767-041-24
; Sequence 24, Application US/09767041
; Patent No. US20020055168A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Hilda
; TITLE OF INVENTION: STREPTOCOCCUS SUIIS VACCINES AND DIAGNOSTIC TESTS
; FILE REFERENCE: 2183-4726
; CURRENT APPLICATION NUMBER: US/09/767,041
; PRIOR FILING DATE: 2001-01-22
; PRIOR APPLICATION NUMBER: PCT/NL99/00460
; PRIOR FILING DATE: 1999-07-19
; PRIOR APPLICATION NUMBER: EP98202465.5
; PRIOR FILING DATE: 1998-07-22
; PRIOR APPLICATION NUMBER: EP98202467.1
; PRIOR FILING DATE: 1998-07-22
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: Patent In version 3.0
; SEQ ID NO 24
; LENGTH: 338
; TYPE: PRT
; ORGANISM: Streptococcus suis
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: CDS2P

```

Thu May 13 11:53:08 2004

us-09-930-440b-6.rapb

Page 3

US-09-767-041-24

Query Match	28.9%	Score 544.5	DB 9	Length 338
Best Local Similarity	36.0%	Pred. No. 2.7e-46		
Matches 123	Conservative 73	Mismatches 129	Indels 17	Gaps 8

Qy	20	FIILAEIGCHQNDLIVAKRMIEMAKESCADAKROKSELEKPKYRKALBEPYKSKSWCK	79
Db	3	YIIAEIGCHNHNDVHIAKRMVEAVANDCGVDANKOETEKADILLISKVAPKAEV--OKITTES	61
Qy	80	TYG--EHKHELEFSDQYRELORYABEYIGFTLASGQDEMAVESEHILNLPFFKVSOGT	137
Db	62	SDSQLEMTRELEISFEERYLDNDYCLLEKGVDFSTPEBDSBLDELSTDMVPYKIRSGEL	121
Qy	138	NRPPELEKTAKKGRPWAVISGMSQMDTYKQVYQIVKPLNPN---FCFLQCTSAYPELOPE	193
Db	122	TNLPYLEKIGROAKKVVILSTGANAND---EIHQAVKILQENGTDTISILCTTEYPPPYA	178
Qy	194	DWNLKRTISEYQKLFPIIPICYSGHETGIALSVAAALGAKYLEHILTLDTWKGSDHSAS	253
Db	179	ALNINLNTLTKREPMLTIGYSDHSVGEVPIALAAAGALIELHGFLLIDMBMSGPDHKAS	228
Qy	254	LRBGEALAEIVRSYALVERALGSPYKOLPCENACNEKL--GKSVYARVKIPEGTILTMQ	311
Db	239	ATPDILALAVKGRIVYEGSLGKEFKB--PEBEVYANKIVAKKSIVAKKALAKGEVTEEN	296
Qy	312	LTVKVGEP--KAYPPEDEIEMLVGKVALVYVEEDGTINEBYND	352
Db	297	ITVK--RPGNGISPMKEVYVGLQVSDQDFEESONICHSAPEN	336

RESULT 5

US-10-156-761-10910
 : Sequence 10910, Application US/10156761
 : Publication No. US20030119018A1
 : GENERAL INFORMATION:
 : APPLICANT: OMURA, SATOSHI
 : APPLICANT: IKEDA, FAEUD
 : APPLICANT: ISHIKAWA, JUN
 : APPLICANT: HORIKAWA, HIROSHI
 : APPLICANT: SHIBA, TADAYOSHI
 : APPLICANT: SAKAKI, YOSHIYUKI
 : APPLICANT: HATTORI, MASAHIRA
 : TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 : FILE REFERENCE: 249-262
 : CURRENT APPLICATION NUMBER: US/10/156, 761
 : CURRENT FILING DATE: 2002-05-29
 : PRIOR APPLICATION NUMBER: JP 2001-204089
 : PRIOR FILING DATE: 2001-05-30
 : PRIOR APPLICATION NUMBER: JP 2001-272697
 : PRIOR FILING DATE: 2001-08-02
 : NUMBER OF SEQ ID NOS: 15109
 : SEQ ID NO 10910
 : LENGTH: 311
 : TYPE: PRT
 : ORGANISM: Streptomyces avermitilis
 : US-10-156-761-10910

[illegible][illegible]

RESULT 6

US-09-930-440B-8
Sequence 8, Application US/09930440B
Patent No. US20020142386A1
GENERAL INFORMATION:
APPLICANT: Betendauhn et al.
TITLE OF INVENTION: Engineering Intracellular Sialylation Pathways
FILE REFERENCE: PFS09P2
CURRENT APPLICATION NUMBER: US/09/930,440B
CURRENT FILING DATE: 2001-08-16
PRIOR APPLICATION NUMBER: 60/227,579
PRIOR FILING DATE: 2000-08-25
PRIOR APPLICATION NUMBER: 09/516,793
PRIOR FILING DATE: 2000-03-01
PRIOR APPLICATION NUMBER: 60/169,624
PRIOR FILING DATE: 1999-12-08
PRIOR APPLICATION NUMBER: 60/122,582
PRIOR FILING DATE: 1999-03-02
NUMBER OF SEQ ID NOS: 8
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 8
LENGTH: 346
TYPE: PRT
ORGANISM: Escherichia coli
US-09-930-440B-8

Query Match	27.1%;	Score 510.5;	DB 9;	Length 346;
Best Local Similarity	36.3%;	Pred. No. 8e-43;		
Matches 123; Conservative	61;	Mismatches 140;	Indels 15;	Gaps

```

0Y 20 F1IAIGONHGOEGLDVAKNINIMAKCGADCAKFOKSEIEFEKNKALERPTSCHSNG- 78
Db 5 Y1VAIEICNHSVDIAREKILKKEAGVNAAYXOFPAKDKLISALPAEYQIKNTGEL 64
0Y 79 KYTGEKRLHEPSHDQYRELQRYAEVGIFFPAISGMDENAVFELHNVFPFKVSSGDTN 13
Db 65 ESQLEMTKKEEMKYDYIHLMEYVAISLNDVSTPEDESDIDFLSKOKIKMISGELL 12
0Y 139 NEPLIETAKGRP---WVSSGWSMDTKQVQV---KPLNDFCLOOTSAYPLOP 19
Db 125 NEPLIETALPLPDKKIISTGMAIDELKQSVSFIINNKVPAGNITTLICNTEYPTPF 18
0Y 193 EDVNLRAVISEYQULFPDIPIGYSGHETGALISVAANLAKLBEHITLDTXMKSGDJA 25
Db 185 EDVNLNAINDLKHPFNNIGFSDHSGFVAALAAAPYIITEKFTLIDKMSGPDHIA 24
0Y 253 SLEPGEALAVRSVRLVERALASPTQÜLLPCENACNEKLGKSVAVAKVIBESTILTMQML 31
Db 245 SIEDELMKHCIGRCVCKSLGNSKVPASERKNITVARKKIIIAKTEILKEGVSEKNI 30
0Y 313 TVKQGEF-KAIPPEDLINLVGKKVYLTVYEDDITIMELV 350
Db 305 TTK---RKGNGISPMENYNLLGK-----IAQÖD-FIPBELI 336

```

RESULT 7

US-10-106-698-6926
; Sequence 6926, Application US/10106698
; Publication No. US20030109690A1

Thu May 13 11:53:08 2004

us-09-930-440b-6.rapb

Page 4

```
GENERAL INFORMATION:
APPLICANT: Ruben et al.
TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
FILE REFERENCE: P4005P1
CURRENT APPLICATION NUMBER: US/10/106,698
CURRENT FILING DATE: 2002-03-27
PRIOR APPLICATION NUMBER: PCT/US00/26524
PRIOR FILING DATE: 2000-09-28
PRIOR APPLICATION NUMBER: US 60/157,137
PRIOR FILING DATE: 1999-09-29
PRIOR APPLICATION NUMBER: US 60/163,280
PRIOR FILING DATE: 1999-11-03
NUMBER OF SEQ ID NOS: 8564
SOFTWARE: PatentIn Ver. 3.0
SEQ ID NO 6926
LENGTH: 123
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (83)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: MISC FEATURE
LOCATION: (87)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: MISC FEATURE
LOCATION: (111)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: MISC FEATURE
LOCATION: (112)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: MISC FEATURE
LOCATION: (121)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-106-698-6926

Query Match      20.2% Score 381; DB 14; Length 123;
Best Local Similarity 93.7%; Pred. No. 2,2e-30;
Matches 74; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Cy 183 OCTSAVPLEDEVDVNLRISEYOKLPDIPIGYSGHETGIAISVAVALGAKVLERHTITLD 242
Db 24 KCTSAIPLOPEDVNLRISEYOKLPDIPIGYSGHETGIAISVAVALGAKVLERHTITLX 83
QY 243 KTWKGSDBSASLEPGEIAB 261
Db 84 KTWKGSDBSASLEPGEIAB 102

RESULT 8
US-09-816-028A-35
Sequence 35; Application US/09816028A
Patent No. US20020042369A1
GENERAL INFORMATION:
APPLICANT: Gilbert, Michel
APPLICANT: Makarchuk, Warren W.
APPLICANT: National Research Council of Canada
TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
FILE REFERENCE: 019633-000111US
CURRENT APPLICATION NUMBER: US/09/816,028A
PRIOR FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: US 60/118,213
PRIOR FILING DATE: 1999-02-01
PRIOR APPLICATION NUMBER: US 09/495,406
PRIOR FILING DATE: 2000-01-31
NUMBER OF SEQ ID NOS: 49
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 35
LENGTH: 346
TYPE: PRT
ORGANISM: Campylobacter jejuni
FEATURE:
```

```
OTHER INFORMATION: sialic acid synthase from C. jejuni OH4384 (ORF 8a)
OTHER INFORMATION: of lipooligosaccharide (LOS) biosynthesis locus)
US-09-816-028A-35

Query Match      17.8% Score 336.5; DB 9; Length 346;
Best Local Similarity 29.1%; Pred. No. 3.7e-25;
Matches 96; Conservative 63; Mismatches 164; Indels 7; Gaps 4;

Cy 21 ITAEIGONHOGDLVAKMIRMAECGADCAKFEKSELEFFKFKKALE-PPYTSKHSWGK 79
Db 19 VPEIGINHNHSLALAKIMVDAISTAKIIKQTHIVEDMSKAKVLPGNAKIS--- 75
QY 80 TYGEHKKHLEPSHDQYRELQRYAEVGIPTASGDMANAEFLHNLVPPFKVSGDTNN 139
Db 76 -IYEIMQKCALDYDDELAKETKELGIVYLSTPSPAGANRLEDMGVSAPKISGECNN 134
QY 140 FPYLEKTKAGRPWVVISGQSMQMDTKOVQIVPLNPFCELOCTSAVPLEDEVDVNLRI 199
Db 135 YPLIKHIAFKKPMIVSTGNMISIEIKPTVKILLDNHLPVLMHTNMLPPTPHNLVRLNA 194
QY 200 ISEYOKLPDIPIGYSGHETGIAISVAVALGAKVLERHTITLDKTWKGSDBSASLEPGEI 259
Db 195 MLEIKKEF-SCWVGILSHHTDNLACLGAVALGACVLRHPTDSWHRSGPDIVCSMDTQAL 253
QY 260 AEL-VRSVRLVERALGSPTRQLLPCEMACNEKLGKSVAVAKIPEGITLITMDMLTVKGE 318
Db 254 KELIIQSQAIMEGNNESKRAAKOEVTTIDPARASVSTIDIKKGEVLMDNIMVWRPG 313
QY 319 PKAYPPEDIFNLVGRKVLVTVSEDDTIMEE 348
Db 314 LQGISAEFENILGKKALRDIENDTOLSYE 343

RESULT 9
US-10-303-161-35
Sequence 35; Application US/10303161
Publication No. US20030148459A1
GENERAL INFORMATION:
APPLICANT: Gilbert, Michel
APPLICANT: Makarchuk, Warren W.
APPLICANT: National Research Council of Canada
TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
FILE REFERENCE: 019633-000111US
CURRENT APPLICATION NUMBER: US/10/303,161
PRIOR FILING DATE: 2002-11-21
PRIOR APPLICATION NUMBER: US/09/816,028
PRIOR FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: US 60/118,213
PRIOR FILING DATE: 1999-02-01
PRIOR APPLICATION NUMBER: US 09/495,406
PRIOR FILING DATE: 2000-01-31
NUMBER OF SEQ ID NOS: 49
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 35
LENGTH: 346
TYPE: PRT
ORGANISM: Campylobacter jejuni
FEATURE:

US-10-303-161-35

Query Match      17.8% Score 336.5; DB 14; Length 346;
Best Local Similarity 29.1%; Pred. No. 3.7e-25;
Matches 96; Conservative 63; Mismatches 164; Indels 7; Gaps 4;

Cy 21 ITAEIGONHOGDLVAKMIRMAECGADCAKFEKSELEFFKFKKALE-PPYTSKHSWGK 79
Db 19 VPEIGINHNHSLALAKIMVDAISTAKIIKQTHIVEDMSKAKVLPGNAKIS--- 75
QY 80 TYGEHKKHLEPSHDQYRELQRYAEVGIPTASGDMANAEFLHNLVPPFKVSGDTNN 139
Db 76 -IYEIMQKCALDYDDELAKETKELGIVYLSTPSPAGANRLEDMGVSAPKISGECNN 134
QY 140 FPYLEKTKAGRPWVVISGQSMQMDTKOVQIVPLNPFCELOCTSAVPLEDEVDVNLRI 199
Db 135 YPLIKHIAFKKPMIVSTGNMISIEIKPTVKILLDNHLPVLMHTNMLPPTPHNLVRLNA 194
QY 200 ISEYOKLPDIPIGYSGHETGIAISVAVALGAKVLERHTITLDKTWKGSDBSASLEPGEI 259
Db 195 MLEIKKEF-SCWVGILSHHTDNLACLGAVALGACVLRHPTDSWHRSGPDIVCSMDTQAL 253
QY 260 AEL-VRSVRLVERALGSPTRQLLPCEMACNEKLGKSVAVAKIPEGITLITMDMLTVKGE 318
Db 254 KELIIQSQAIMEGNNESKRAAKOEVTTIDPARASVSTIDIKKGEVLMDNIMVWRPG 313
QY 319 PKAYPPEDIFNLVGRKVLVTVSEDDTIMEE 348
Db 314 LQGISAEFENILGKKALRDIENDTOLSYE 343
```

Db 76 -IYEWKCALDYKDELALKEYTEKGLVYLSTPFSRAGANRLDMDGVSAPKIGSGECNN 134
 Qy 140 FPLLEKTKAGGRPMWISSGQSMQMDTKQVQIVKPLNPNCFLOQSAVPLQPEDVNLAV 199
 Db 135 YPLIKHIAFKKPMIVSTGNISIESIKPTVKILDNBIPVLMTNLTPTPHNLVRLNA 194
 Qy 200 ISEYOKLFPDIPIGSGHETGIAISYAVALGAKVLERHITLDKTKWGSDBSASLEPGEI 259
 Db 195 MELKKERF-SCHWGLSDHTDNLACGAVNLGACVLERHPTDMSHRSRGPDIIVCSMDTOAL 253
 Qy 260 AEL-VRSVRLVERALSSPTKQLLPCEMACNEKIGKSVAVAKVPEGTILTMDLTVYVGE 318
 Db 254 KELIISSEQWAINRGNNESKKAQKQOVITIDFAPASVSIKDIKKEVLSMDNIWVRPG 313
 Qy 319 PKAYPPEDIFNLGKRVLTVEEDDTIMEE 348
 Db 314 LGGISAAPFENILGKRLRDIENDTQLSYE 343

RESULT 10
 US-10-303-118-35
 ; Sequence 35, Application US/10303118
 ; Publication No. US20030157655A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Gilbert, Michel
 ; APPLICANT: Makarchuk, Warren W.
 ; APPLICANT: National Research Council of Canada
 ; TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
 ; TITLE OF INVENTION: Gangliosides and Ganglioside Mimics
 ; FILE REFERENCE: 019633-000111US
 ; CURRENT APPLICATION NUMBER: US/10/303,118
 ; PRIOR FILING DATE: 2002-11-21
 ; PRIOR APPLICATION NUMBER: US/09/816,028
 ; PRIOR FILING DATE: 2001-03-21
 ; PRIOR APPLICATION NUMBER: US 60/118,213
 ; PRIOR FILING DATE: 1999-02-01
 ; PRIOR APPLICATION NUMBER: US 09/495,406
 ; PRIOR FILING DATE: 2000-01-31
 ; NUMBER OF SEQ ID NOS: 49
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 35
 ; LENGTH: 346
 ; TYPE: PRF
 ; ORGANISM: Campylobacter jejuni
 ; FEATURE:
 ; OTHER INFORMATION: sialic acid synthase from C. jejuni OH4384 (ORF 84)
 ; OTHER INFORMATION: of lipooligosaccharide (LOS) biosynthesis locus)
 US-10-303-118-35

Query Match 17.8%; Score 336.5; DB 14; Length 346;
 Best Local Similarity 29.1%; Pred. No. 3.7e-25;
 Matches 96; Conservative 63; Mismatches 164; Indels 7; Gaps 4;

Qy 21 IIAEIGNHOGDLDVAKKMRMAKCGADCAKFKXSELEFFKRLAE-RPYTSKSHWGK 79
 Db 19 VPEIGINHGSLKAKIMVDAAFSTGAKIIGHQTHIVEDMSYAAKXVIFGNKXIS--- 75
 Qy 80 TYGHNKRLFSHOYRELQRYAEVGIFFETASGMDENAVEFLHNLNVPFKVSGSDTN 139
 Db 76 -IYEWKCALDYKDELALKEYTEKGLVYLSTPFSRAGANRLDMDGVSAPKIGSGECNN 134
 Qy 140 FPLLEKTKAGGRPMWISSGQSMQMDTKQVQIVKPLNPNCFLOQSAVPLQPEDVNLAV 199
 Db 135 YPLIKHIAFKKPMIVSTGNISIESIKPTVKILDNBIPVLMTNLTPTPHNLVRLNA 194
 Qy 200 ISEYOKLFPDIPIGSGHETGIAISYAVALGAKVLERHITLDKTKWGSDBSASLEPGEI 259
 Db 195 MELKKERF-SCHWGLSDHTDNLACGAVNLGACVLERHPTDMSHRSRGPDIIVCSMDTOAL 253
 Qy 260 AEL-VRSVRLVERALSSPTKQLLPCEMACNEKIGKSVAVAKVPEGTILTMDLTVYVGE 318
 Db 254 KELIISSEQWAINRGNNESKKAQKQOVITIDFAPASVSIKDIKKEVLSMDNIWVRPG 313

Qy 319 PKAYPPEDIFNLGKRVLTVEEDDTIMEE 348
 Db 314 LGGISAAPFENILGKRLRDIENDTQLSYE 343

RESULT 11
 US-10-303-128-35
 ; Sequence 35, Application US/10303128
 ; Publication No. US20030157656A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Gilbert, Michel
 ; APPLICANT: Makarchuk, Warren W.
 ; APPLICANT: National Research Council of Canada
 ; TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
 ; TITLE OF INVENTION: Gangliosides and Ganglioside Mimics
 ; FILE REFERENCE: 019633-000111US
 ; CURRENT APPLICATION NUMBER: US/10/303,128
 ; PRIOR FILING DATE: 2002-11-21
 ; PRIOR APPLICATION NUMBER: US/09/816,028
 ; PRIOR FILING DATE: 2001-03-21
 ; PRIOR APPLICATION NUMBER: US 60/118,213
 ; PRIOR FILING DATE: 1999-02-01
 ; PRIOR APPLICATION NUMBER: US 09/495,406
 ; PRIOR FILING DATE: 2000-01-31
 ; NUMBER OF SEQ ID NOS: 49
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 35
 ; LENGTH: 346
 ; TYPE: PRF
 ; ORGANISM: Campylobacter jejuni
 ; FEATURE:
 ; OTHER INFORMATION: sialic acid synthase from C. jejuni OH4384 (ORF 84)
 ; OTHER INFORMATION: of lipooligosaccharide (LOS) biosynthesis locus)
 US-10-303-128-35

Query Match 17.8%; Score 336.5; DB 14; Length 346;
 Best Local Similarity 29.1%; Pred. No. 3.7e-25;
 Matches 96; Conservative 63; Mismatches 164; Indels 7; Gaps 4;

Qy 21 IIAEIGNHOGDLDVAKKMRMAKCGADCAKFKXSELEFFKRLAE-RPYTSKSHWGK 79
 Db 19 VPEIGINHGSLKAKIMVDAAFSTGAKIIGHQTHIVEDMSYAAKXVIFGNKXIS--- 75
 Qy 80 TYGHNKRLFSHOYRELQRYAEVGIFFETASGMDENAVEFLHNLNVPFKVSGSDTN 139
 Db 76 -IYEWKCALDYKDELALKEYTEKGLVYLSTPFSRAGANRLDMDGVSAPKIGSGECNN 134
 Qy 140 FPLLEKTKAGGRPMWISSGQSMQMDTKQVQIVKPLNPNCFLOQSAVPLQPEDVNLAV 199
 Db 135 YPLIKHIAFKKPMIVSTGNISIESIKPTVKILDNBIPVLMTNLTPTPHNLVRLNA 194
 Qy 200 ISEYOKLFPDIPIGSGHETGIAISYAVALGAKVLERHITLDKTKWGSDBSASLEPGEI 259
 Db 195 MELKKERF-SCHWGLSDHTDNLACGAVNLGACVLERHPTDMSHRSRGPDIIVCSMDTOAL 253
 Qy 260 AEL-VRSVRLVERALSSPTKQLLPCEMACNEKIGKSVAVAKVPEGTILTMDLTVYVGE 318
 Db 254 KELIISSEQWAINRGNNESKKAQKQOVITIDFAPASVSIKDIKKEVLSMDNIWVRPG 313

Qy 319 PKAYPPEDIFNLGKRVLTVEEDDTIMEE 348
 Db 314 LGGISAAPFENILGKRLRDIENDTQLSYE 343

RESULT 12
 US-10-303-134-35
 ; Sequence 35, Application US/10303134
 ; Publication No. US20030157657A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Gilbert, Michel
 ; APPLICANT: Makarchuk, Warren W.
 ; APPLICANT: National Research Council of Canada
 ; TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of

Query Match	17.8%;	Score 336.5;	DB 14;	Length 346;
Best Local Similarity	29.1%;	Pred. No. 3.7e-25;		
Matches	96;	Conservative	63;	Mismatches 164;
			Indels	7;
			Gaps	4

RESULT 13

LENGTH: 346

Query Match	17.8%;	Score 336.5;	DB 14;	Length 346;
Best Local Similarity	29.1%;	Pred. No. 3.7e-25;		
Matches 96;	Conservative 63;	Mismatches 164;	Indels 7;	Gaps 4

US-09-882

Query Match	16.1%	Score 303;	DB 10;	Length 340;
Best Local Similarity	27.7%	Pred. No. 9.1e-22;		
		57.0% matches	149;	
			TrdeIs 22;	Gans 10

D

```

18 PCPIIAETGONHODLVAKMIRAKKEGACCAAPRK-----SELEFKRPALESPY 72
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
4  PPKVIAEASANNODLAKESLHAKESGAPVYLQYTPSCMTLNSXEDPFIQSTLW 63
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
73 SKSHWAKTYGHEKHLFESHOYRELQYAEVGVFTTASGMDENKAEVPEHLENVPEVKY 133
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
64 KRENIVELYQKASTPLEM--HEHLEFLAKYLDGLG--SSPESQALLESLSLNCMPYKI 119
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```


Thu May 13 11:53:08 2004

us-09-930-440b-6.rapb

Page 7

QY 133 GSGDINNPFYLEKTAAGKRPMTVSSGMSMDTKQVYQVYKPINPNF--CELOCTSAVPL 130
DB 120 ASFEIVDLIDLEKARQKPIILSSGIAHTELODAISLCRVN-NFDITLTKCVSAVPS 178
QY 191 QPBDVNLRTVSEYOKLPDIPITIGYSGHETGIAISVAALGAKVLEHRTILDITKWSGDH 250
DB 179 KIBDANLISVVKGEIF-GVYFEGSDHTIGSLCPILATTTGASWIEKHFIANKSLQTPDS 237
QY 251 SASIEPELAEIVRSVLRALGSPKOLLPCGMAGNEKLGKSVAKVXIPGTTILT-- 308
DB 238 AFSDFNGFKSMWEALQSVLALGEEPRINPKTLERRPPASLFFVTKIOQGEALTEN 237
QY 309 -MDMLTVKVG-EPKAYPPEDIFNLVGRK 334
DB 298 NIKALRPNLGLHPRFKYK-----EILGQK 320

RESULT 15
US-10-335-977-7626
Sequence 7626, Application US/10335977
Publication No. US20040052799A1
GENERAL INFORMATION:

APPLICANT: DOUGLAS SMITH et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
RELATING TO HELICOBACTER PYLORI FOR
DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 10031

CORRESPONDENCE ADDRESS:

ADDRESSEE: LAHIVE & COCKFIELD

STREET: 28 State Street

CITY: Boston

STATE: Massachusetts

COUNTRY: USA

ZIP: 02109-1875

COMPUTER READABLE FORM:

MEDIUM TYPE: CD-ROM ISO9660

COMPUTER: IBM PC Compatible

OPERATING SYSTEM: Windows NT 4.0

SOFTWARE: UNIX

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/335,977

FILING DATE: 30-Dec-2002

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/993,002

FILING DATE: 17-DEC-1997

ATTORNEY/AGENT INFORMATION:

NAME: Mandragouras, Amy E.

REGISTRATION NUMBER: 36,207

REFERENCE/DOCKET NUMBER: GTN-018

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617)227-7400

TELEFAX: (617)742-4214

INFORMATION FOR SEQ ID NO: 7626:

SEQUENCE CHARACTERISTICS:

LENGTH: 340 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULAR TYPE: protein

HYPOTHEICAL: YES

ORIGINAL SOURCE:

ORGANISM: Helicobacter pylori

FEATURE:

NAME/KEY: misc feature

LOCATION: (B) LOCATION 1...340

SEQUENCE DESCRIPTION: SEQ ID NO: 7626:

US-10-335-977-7626

QY 21 IIAIGQHODLVAKKIMAKGCAKAFK-----SELEFKFRKALERTPYTSKH 75

DB 7 IVALISANENODLNLAKESLHAIKESGADPVKIQVTPSCMTLDSK-----EDPFILOG 60
QY 76 S-WGR--TYGHEKR--HLEPSHDQYRELFQYAEVGFPTASGMDMAVEFLHELVNPF 129
DB 61 TLMDKENLYGLYQKASTPLEW-HAELFELAKL-DLGIF--SPPSKALELLESIDCTM 116
QY 130 FKVSGDINNPFYLEKTAAGKRPMTVSSGMSMDTKQVYQVYKPINPNF--CELOCTSA 187
DB 117 KIASFEIVDLIDLEKARQKPIILSSGIAHTELODAISLCRVN-NFDITLTKCVSA 175
QY 188 YPLQPDVNLRTVSEYOKLPDIPITIGYSGHETGIAISVAALGAKVLEHRTILDITKWSGDH 247
DB 176 YPEKIBDAHLISVVKGEIF-GVYFEGSDHTIGSLCPILATTTGASWIEKHFIANKSLQTPDS 234
QY 248 SDSASLEKGELEIVRSVLRALGSPKOLLPCGMAGNEKLGKSVAKVXIPGTTILT 307
DB 235 PDSAFMDTFNGFMSVGAIKQSVLALGEEPRINPKTLERRPPASLFFVTKIOQGEALTEN 234
QY 308 TMD--MLTVKVG-EPKAYPPEDIFNLVGRK 334
DB 295 TSDNIRALRPNLGLHPRFKYK-----EILGQK 320

Search completed: May 6, 2004, 09:25:07
Job time: 40.8108 secs

Query Match 15.2%; Score 287; DB 12; Length 340;
Best Local Similarity 27.8%; Pred. No. 3.8e-20;
Matches 92; Conservative 68; Mismatches 137; Indels 34; Gaps 14;

Thu May 13 11:53:11 2004

us-09-930-440b-7.rge

Page 1

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2004, 07:15:34 ; Search time 4297.94 Seconds

(without alignments)
10679.608 Million cell updates/sec

Title: US-09-930-440b-7

Perfect score: 1059

Sequence: 1 atgagtaataatataatcgt.....aatgagaacaaatattatg 1059

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 3470272 segs, 2167151695 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenBank :
1: gb_da :
2: gb_hg :
3: gb_in :
4: gb_om :
5: gb_ov :
6: gb_pat :
7: gb_ph :
8: gb_pl :
9: gb_pr :
10: gb_ro :
11: gb_sts :
12: gb_sy :
13: gb_un :
14: gb_vi :
15: gb_wa :
16: em_fun :
17: em_hum :
18: em_in :
19: em_om :
20: em_ov :
21: em_or :
22: em_ov :
23: em_pat :
24: em_ph :
25: em_pl :
26: em_ro :
27: em_sts :
28: em_un :
29: em_vi :
30: em_hg_hum :
31: em_hg_in :
32: em_hg_other :
33: em_hg_mus :
34: em_hg_pln :
35: em_hg_rtd :
36: em_hg_mam :
37: em_hg_vit :
38: em_sy :
39: em_hgo_hum :
40: em_hgo_mus :
41: em_hgo_other :

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1059	100.0	1781	1	ECU05248
2	520	49.1	12105	1	AF361371 Escherich
3	328.6	31.1	17596	1	AF355776 Streptoco
4	328	31.0	16448	1	AF337958 Streptoco
5	326.4	30.8	17276	1	AF163833 Streptoco
6	324.8	30.7	10206	1	AB017355 Streptoco
7	324.8	30.7	18239	1	AF349539 Streptoco
8	324.8	30.7	21365	1	AE014245 Streptoco
9	324.8	30.7	25021	1	AB028896 Streptoco
10	324.8	30.7	95596	6	AX602206 Streptoco
11	324.8	30.7	128050	1	SA676849 Streptoco
12	321.4	30.3	1026	6	AX607023 Streptoco
13	312	29.5	9987	1	AB050723 Streptoco
14	300.6	28.4	26335	1	CJ11168X5 Streptoco
15	294.2	27.8	13390	1	AY102621 Streptoco
16	275.4	26.0	11938	1	AE011338 Streptoco
17	214.4	20.2	1849	1	AF195054 Streptoco
18	210.8	19.9	300171	1	AE015942 Streptoco
19	200.4	18.9	301235	1	AE016799 Streptoco
20	195	18.4	349562	1	BX569690 Streptoco
21	182.4	17.2	32661	1	LPW7311 Streptoco
22	170.6	16.1	12858	1	AE007719 Streptoco
23	166.2	15.7	14561	1	U67549 Streptoco
24	166.2	15.7	110000	6	AE271569_04 Streptoco
25	137.4	13.0	300254	1	AE016933 Streptoco
26	131.6	12.4	3876	1	AY034084 Streptoco
27	130.4	12.3	300000	1	AP005073 Streptoco
28	122.6	11.6	1388	1	CJA00855 Streptoco
29	118	11.1	3071	1	AF195055 Streptoco
30	118	11.1	4668	1	AF257460 Streptoco
31	116.4	11.0	13484	1	AF400047 Streptoco
32	116.4	11.0	13484	1	AY044156 Streptoco
33	116.4	11.0	282183	1	CJ11168X4 Streptoco
34	114.8	10.8	11474	1	AF215659 Streptoco
35	114	10.8	11442	1	AF400048 Streptoco
36	113.2	10.7	11455	1	AY044868 Streptoco
37	110	10.4	4749	1	AF305571 Streptoco
38	110	10.4	11474	1	AF167344 Streptoco
39	109	10.3	10292	1	AE011339 Streptoco
40	108.4	10.2	12390	1	AF401528 Streptoco
41	108	10.2	5067	1	AF126256 Streptoco
42	107.6	10.2	1050	1	NM040740 Streptoco
43	107.6	10.2	4176	1	NM5053 Streptoco
44	107.6	10.2	4908	1	NM5053 Streptoco
45	107.6	10.2	15668	1	AE002366 Streptoco

ALIGNMENTS

RESULT 1
ECU05248
LOCUS ECU05248
DEFINITION Escherichia coli polysialic acid gene cluster region 2 (neud and
neub) genes, complete cds.
ACCESSION U05248
VERSION U05248.1 GI:454079
KEYWORDS
SOURCE Escherichia coli
ORGANISM Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.

REFERENCE
AUTHORS Annunziato, P.W., Wright, L.F., Vann, W.F. and Silver, R.P.
TITLE Nucleotide sequence and genetic analysis of the neod and Neub genes

Page 2

in region 2 of the polysialic acid gene cluster of *Escherichia coli* K1
 U. Bacteriol. 177 (2), 312-319 (1995)
 95113767
 7814319
 2 (bases 1 to 1781)
 Silver, R.P.
 Direct Submission
 Submitted (19-JAN-1994) Richard P. Silver, Microbiology and
 Immunology, University of Rochester, 601 Elmwood Ave, Rochester, NY
 14642, USA
 e-mail: rpsilver@u.rochester.edu

```
/organism="Escherichia coli"  
/mol_type="genomic DNA"
```

```

/codon_start=1
/gene="neuN"

```

```

/gene="neuB"
/codon_start=1

```

```

/protein_id="AA043302.1"
/db_xref="GI:454081"
/translation="MSNVIYASICGHHGSDVIAEMILKAKEAIVNNAKQETFEKAA
KLIASIAKAYOIKNGEISGCEFMKEMKYDDLHMEVAISINDVSTEDPSTK
DIDELPAIAIKKIKTIEPCELLNIPYELAKIAIPDKKIITSGAATIDELKQASV
EINAKYPAIGNTILAHGNEVSTPEEDINAAINDIKKHPKNNIGDHSIDSEYAAII
AVAPGITELEHFTLIDMSGDELLAIDEDELKLCISYRCIERSIGNSGVYVYIA
KNNKIVAKRSIIIAKTEIKGEVSEKNTTKRNGISPMENINLLAKIAEDTIPDD
LIHSEEPKQSGS"

```

1 ATGAGTATATATATATATCGTTCGCTGAATTGGTTCACACCATTAATGCTAGCTGTGATATT 60

[illegible][illegible]

and Reeves, P.R.
Direct Submission
Journal Submitted (14-MAR-2001) Department of Microbiology, The University
of Sydney, Sydney, NSW 2006, Australia
Location/Qualifiers

FEATURES

1. 12105
/organism="Escherichia coli"
/mol_type="genomic DNA"
/db_xref="taxon:562"
/note="serogroup: O104"
141..737
/gene="galP"
141..737
/gene="galP"
/codon_start=1
/transl_table=1
/product="glucose-1-phosphate uridylyltransferase"
/protein_id="AAK64366.1"
/db_xref="gi:14517801"
/translation="MNVKQGEPLGSGILCAPALGDPVYVYLPVVIDDAADPL
KYNLMTARFNEFGRSOVLAKMPGLSESVIQTEPLDRGKSRIVELEKPDQ
PQLDSIDIMAVGVYLSADIMPELEKTOPGWSIQTLDAIELAKQSDVADMLMTGD
SYDCGKRWYMAQFYKYGIRNLKEGAKFKGIKELISE"
1098..1718
/gene="wcdX"
1098..1718
/gene="wcdX"
/codon_start=1
/transl_table=1
/product="wcdX"
/protein_id="AAK64367.1"
/db_xref="gi:14517802"
/translation="MKKKLIIIGAGFAKAVIDLDHBYALGFPVDSKSGHQGP
IVGHSLSPISTPTVYFIAIGDPDRALMISOIDLKLTIVVIDKALVSTRSSIG
TCIYIGKATYCDSDLEDGVYNTPLALIEGHNISYCTNISTNVVLNGVYIGERSF
IGSTVNGQIKIGSSTIVSGSVYIRIPIRVVAGTPTRLIRER"
1723..2763
/gene="nab"
1723..2763
/gene="nab"
/note="NeuAc condensating enzyme"
/codon_start=1
/transl_table=1
/product="nab"
/protein_id="AAK64368.1"
/db_xref="gi:14517803"
/translation="MSKIYIYABIGCHNGDFNLAKKQVEAKKAGVDAVKPQTFRAD
QLISKYAPAEYQIKVTGNSQDLETRKLEPYBEIKLESYAKELGLDVPETPDF
ESIDFLSRNOKMKRIPSGELNLPYEKISRLPISEKRIIVLSTGMATIBETLAAVLA
LTKKMAPEKTIICHTYEPTPEYDNLNLSIIGICRPGKQVGPDSHSTPAAVLA
SVYGIPIRIGHTLAKGFRSGPBKASVTPDELALCEGIRAYEKLCSAKIYVTSB
OKKIVAKSIYAKNIRKSEIYITIDITTRKPNGISPMWYDILGKAEKRFDEDO
LISHSFDEDEI"
2763..4034
/gene="nabC"
2763..4034
/gene="nabC"
/note="CMP-NeuNAc synthetase"
/codon_start=1
/transl_table=1
/product="nabC"
/protein_id="AAK64369.1"
/db_xref="gi:14517804"
/translation="MTRIAIIPAPAGSKGIPKRVNMLIDKPLIATVIEAASQCF
NKIIVTSDSEYKSIARQAGAVIIRSEELASPTATVYIENVLSHNDQCPYMLDQ
PTSPRNARHITKEAITLPBGVPRANRPLVSMSTRESISDLMKTLDSNRISTRNNYSG
YRQDNQKYSFPRCAIPICTYBEETLKQHPGADSLATYAKKADSIDIDRIDPEPAI
SRRLQNNKIKDEAVARIRKQFSTHAIKPTLLGHSTFDNWEISDLAGKSVNIG
VSGITSDYNNLIDNARIKEIGDVPALGNDIVYIQGDENDTLRMWKIIDRIYK
LAPNVIIYFLSLPVGNGRVDRNNKVIHEHNAFLPKGLDKINHVRVSLSDMPYDDEGN
LSADPTVQGLHFNKAYROLKEIMELIK"
4031..5194
/gene="nabA"

CDS

4031..5194
/gene="nabA"
/note="GlcNAc-2-epimerase"
/codon_start=1
/transl_table=1
/product="nabA"
/protein_id="AAK64370.1"
/db_xref="gi:14517805"
/translation="MKKILYVGSRAEYIMKRLLOOLKEDSDISLIVATGMECDK
YGMFKYIENDGPRVDKIDIEINNDNNASIKTMSRCLESGIYFQDNKPDAMVILG
DRYEMSYALAAAGNLPITPHGSEKTLGNDYDFIRHSTKMSLILASTHEPKRI
IOMGHPNVYDIGNLGNLSLILNPKDVBQFGLSKYFVVFHFBETLSGVP
BIQVSLKSLHATFKITNFVFGSDSDGSDKTQKIKQPCENKNTKVFVSLIBBY
LALCKVSGQLIGNSSGLIEVPSLGVPTVNGRQKRGVAGATVINCSTEGISEAI
KCLDEHYINGIKNVNPYVYKNNANAYKHKEIFLISNHOYKDFYLSFEL"
5279..6229
/gene="wbaA"
5279..6229
/gene="wbaA"
/note="putative glycosyl transferase"
/codon_start=1
/transl_table=1
/product="wbaA"
/protein_id="AAK64371.1"
/db_xref="gi:14517806"
/translation="MKRMNNVVPDPSYCIIYGLIRNHOADPTIYSDNVSLRDIN
PKKERYIIRKGNNTSKITFYFCEPLRLINPKLISNRDQYRELGDHLEFSK
FLNEFTLEDELAVNRYPHYSLKYLIIQHTPERSRKILSGMIDILDPITNK
VVEFDLIAWELKNSVQKELINLFINYCEDELAVADWILAQPLSEQGPSENKIL
YDBIKEYDDKIVIRQHPRELTDYSYFQKRVNKKAPVLELVVANSPIKAVTLF
SGGIFNICEKRIFKGTSFSLSKHGNIINKEEVAKK"
6226..7338
/gene="wzy"
6226..7338
/gene="wzy"
/note="No antigen polymerase"
/codon_start=1
/transl_table=1
/product="wzy"
/protein_id="AAK64372.1"
/db_xref="gi:14517807"
/translation="MTSYFTINFYVLAIFAVYHERINLFTSGNFKLSIGCVFGL
FLSLKRFVGPDEYTYVYIKNIVAKRIDITTYRGLSTLFSGLDGYLVRYV
YSPLTFEPVFRVILBKIPMGIIIVCFVSGYLDLSIRQVLAVALFVTSIEK
BLRPFYAIILIASFFHLSALMLPVPYVNERKEPPLILIGRLVLYSGFTTEIQ
ALIVNNIYVNEIYGNQSPALHOGFGTGFAPFHLIEFISYNEKONVILNRLFT
AGLMIILAPGNLIMRFGQFYISLILALPLAKSLKEXSYNLLKCLFCLLTFE
OOSGRNYSYOSIFSDFNSLSTPSSDRITINH"
7344..8567
/gene="wbaB"
7344..8567
/gene="wbaB"
/note="putative glycosyl transferase"
/codon_start=1
/transl_table=1
/product="wbaB"
/protein_id="AAK64373.1"
/db_xref="gi:14517808"
/translation="MKNIILPTNNALSTGCARVILNAKSLRNRGHVALLDPRVDNI
HYKIDGIEIYVRDKIRKIVNSPTLNKNSQAFPTNQSVEFNFILYAKKIKSFP
YELAPCEMYAFKKFINTKKEAIINNNAVNDRIYBSKLSNNYVFNHNSPELIF
SRREPSLPLPSKIFRNVKLSVSKGLADILLKGNDRBEIITIVNPFTTELEGS
RIQCEGNIPLKKEIIVVSTLTDKRVDRIRKAPKIRISGEIDILLIGEGQLENDL
ONLKEANIETEKYVFLGQNPNPYFIKADLLISSDSBGLPYIILISLITGPVLT
DCLPGEHLLIWMGDSSLVSLTTSDBNNI ONELAEKAIILCKKDKKVKKSDQ
RPEKKSISWEILLF"
8549..9844
/gene="wzx"
8549..9844
/gene="wzx"
/note="O antigen flippase"
/codon_start=1
/transl_table=1

```

/product="Nzx"
/protein_id="AAK64374.1"
/db_xref="GI:14517803"
/translation="MGDIIIMNMLKYIPILSLRMVTVISFLFSWIVARLFGAS
AGTIPFTYITTTTPTWASGAEIGIAKAVARLAKNEKQIQRVSVLRTRMPLI
SPAEFLYLIEGIEINLILPVMLILGYCEPLWNILISFPGIGITMMWISORAI
FNLFSFLIAFLWGVFLBKCSFPLIKQSPVITITGLISRELIVHRTLNKSP
LYKREBFDKSCCKSLFKIQLOLITMVSQIINIPSVKADIAEIVAKISLILGFE
VLASGVISQVSRVSLKDFSIHKRSVHVSFLGVPVAILILVSKSLILFG
NDFQYCAVLIVLIVSQVFNCLTGACDIIILMFDGEXEKXKQVLTGFEALVLSFII
PYSALGAITAFSAIIVNILDVYSIKKXGFMFKTV"
9869..10675
/gene="wbwc"
/contig="wbwc"
9869..10675
/gene="wbwc"
/contig="wbwc"
/feature="putative glycosyl transferase"
/codon_start=1
/translation="MKPSVLISLYKESALFLHDCBSIASNTCAPOIVIVPBGYIS
/protein_id="AAK64375.1"
/db_xref="GI:14517810"
/translation="MKPSVLISLYKESALFLHDCBSIASNTCAPOIVIVPBGYIS
DOLHVNVEFSLRLPIDIVRLSNVGLAKALNKGQYKNEILFPMDDIDILBERTA
LDSYKSHSPVILVLSALIEEDNTKIRQGRFSVYIHEVILKFKRKNFENFV
PKSVIEKLGQYQHLYMBDYRLRLIAADRTENLSEBIVLVNVAKKMCRKRGYS
YKSEIILAKIKYELQLNLNLTGVTGIRIIRILPVLKLVNILRK"
10796..112105
/gene="gnd"
/contig="gnd"
10796..112105
/gene="gnd"
/contig="gnd"
/feature="6-phosphogluconate dehydrogenase"
/codon_start=1
/translation="Gnd"
/product="Gnd"
/protein_id="AAK64376.1"

```

Query Match 49.1%; Score 520; DB 1; Length 12105;
Best Local Similarity 58.3%; Pred. No. 2.3e-73;
Matches 721; Conservative 0; Mismatches 335; Indels 0; Gaps 0;

```

Db 2203 CTTTCTAAAAATGGAATGGCTCTCTGAAAAGATGACAAATATTACATGTTAAGCATAT 2262
Qy 541 CCAAGCCCTTTGAGATGTAAACCTTAATGCTATTATGATTTGAAAAACACTTCCT 600
Db 2263 CTTACACCATATGAGATGTAAATTTAACTCAATATTTGGTCTAAAGAAATGATTTAAA 2322
Qy 601 AAGAAATACAGGCTCTCTGATCAATCTAGGCGGTTTATGACGATATGCGCGGTG 660
Db 2323 GGAATATACGTTGGCTTTCTGACATTCACCGAGTTATTTATGACGATATGATCAGTT 2382
Qy 661 CTTATGAAATTAATTTATGAAAAACATTTCACTTAATTAATCTATGTCGCCCA 720
Db 2383 CCGATGGAATTAATCAATTCAGAAAAACATTTTACTTAATTAATTTTGAAGGCCCT 2442
Qy 721 GATCATTTGGCTTCAATAGAACCTGATGAACGAAACATCTTGTATTTGGGTCAGGTG 780
Db 2443 GATCATTAAGCATCAATTAACCAAGATGAGCTGATTAATTTGTGTAGCGGATTCGGCT 2502
Qy 781 GTTGAATAATCTTTAGGTTCAATATGATGAAGTGTTCACGCTTCAGAAAGAGATPAA 840
Db 2503 GTTGAATAATCTTTAGGATGATGCAAAATATGTAACAATACTGAACAAAAATPAG 2562
Qy 841 ATCTAGCAAGAAAGTCTATTATGCTTAAACAGATTAATAAAGGTGAGTTTTTCA 900
Db 2563 ATAGTACCAAGAAAGTCAATATGTCGAAATATGCAACATTAATAAAGGTGAATATACG 2622
Qy 901 GAAAAAATATTAACAACAAAAAGACTGTATATGATGATCGATCGATGAGTATAT 960
Db 2623 ATGATATATATCAACCAAGAGCGCCAGAAACGATTTGTCTATGATTTGTATGAT 2682
Qy 961 TTATGCTTAAATATGCAAGCAAGACTTATTCCATGATTAATTAATTCATAGGAA 1020
Db 2683 ATCTTACCAAGAAAGCGGAAAAAATTTGATGATGATCAATTAATTAAGTCAATTCGA 2742
Qy 1021 TTCAAAATATCCGGGGAATATAGAACAAAAATTA 1056
Db 2743 TTCAATGAAACGAGATCTTAATTTAATGACTAATA 2778

```

RESULT 3

AF355776

LOCUS AF355776 17596 bp DNA linear BCT 29-Apr-2001

DEFINITION Streptococcus agalactiae CMC 1/82 type IV capsular polysaccharide

ACCESSION AF355776

VERSION AF355776.1 GI:13876769

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLES

JOURNAL

SUBMITTED (03-MAR-2001)

Department of Infectious Disease,

Immunology and Rheumatology, Children's Hospital and Regional

Medical Center, 4800 Sand Point Way NE, Seattle, WA 98105, USA

Location/Qualifiers

1..17596

/organism="Streptococcus agalactiae"

/mol_type="genomic DNA"

/strain="CMC 1/82"

/db_xref="taxon:1311"

817..17187

/gene="type IV capsular polysaccharide synthesis operon"

817..822

/gene="type IV capsular polysaccharide synthesis operon"

834..839

/gene="type IV capsular polysaccharide synthesis operon"

928..2396

/gene="cpsA"

928..933

RBS

[illegible]

/gene="cpsB"
 /codon_start=1
 /transl_table=11
 /product="glucose-1-phosphate transferase CpsB"
 /protein_id="AAK43608.1"
 /db_xref="GI:13876774"
 /translation="MIQIVVYFSSALITLTLIPNFKSKDILFVLLIHIVTVYISDF
 YRDEMSRSLYEFQVWLKYSFVYIFISSLFFIFKNSPTTLLSPFETIMNSILTL
 LNSFLKRYKRSYAKFSDTFKVLITKSDLSKMTFNKDYHNYIANCLIDSEKDY
 DLKNSRLITKDALTESEFCLIVDOAFINIPILFQYQIDIIINDIEMAGYIVANN
 VEAISFQNGKQIOTSEEGSVIYSMKPKVSHLAKRFDLTGALIGLIGIYVNI
 FLVQIRKDDCPAIFSQNRVSRNRIRIFPKFSBMRPAIDKQDILVHQMQLMKR
 LFDPRIRIKKIFKTKTISIDLPQIVLKGDSIVSTGRTPEDEHKNSVGTGRSL
 FEPGLTGMQISGNNITDPDEIVKLVDYINEMSIWDIKIILLTLKVALTGKAK"
 5958..6417
 /gene="cpsF"
 5958..5963
 /gene="cpsF"
 5968..6417
 /gene="cpsF"
 /codon_start=1
 /transl_table=11
 /product="beta-1,4-galactosyltransferase enhancer CpsF"
 /protein_id="AAK43607.1"
 /db_xref="GI:13876775"
 /translation="MKICLVSSSGGHLAHNLKPIEKEDRFWTEEDKDAQSLTE
 EIVHEFPPNINRWNLKNTLLAKYTLKRESPVLISSGAVAVERPFIYGLRGCT
 VYIEVFDRIIDPRTLTGLVYFVTDKFIIVQWEMKKVYPKALINGGIF"
 6402..6508
 /gene="cpsIVg"
 6402..6407
 /gene="cpsIVg"
 6417..6508
 /gene="cpsIVg"
 /codon_start=1
 /transl_table=11
 /product="beta-1,4-galactosyltransferase CpsIVg"
 /protein_id="AAK43608.1"
 /db_xref="GI:13876776"
 /translation="MIPTVATGHEQQRNRLKEVDRLKGTDAIDQVAFITQGSDEP
 QNOMSKSLSDYDNMSYKKAELIVITHGGPATFNNAVSKKQTIIVRQDQPEHVNK
 HOVDVNVKVMYNDIVVDIERLQNVVYEGTMNRPTLETNRSNFIIEPVTILLKELCD
 ENQ"
 6882..7992
 /gene="cpsIVH"
 6882..6887
 /gene="cpsIVH"
 6895..7992
 /gene="cpsIVH"
 /codon_start=1
 /transl_table=11
 /product="putative polysaccharide repeating unit
 polymerase CpsIVH"
 /protein_id="AAK43609.1"
 /db_xref="GI:13876777"
 /translation="MKIKNSLFPYALIFLVNPFKSLGHBGNSYTKIWMVYALFGSH
 KELDLSYEFBRKVIITFLFIATILINDLFVHKVTFITLIFLALDKDISLKAQFSIL
 IGSRLGTLANQIPKIDLIRIXYINFPDQPLASDLAGHPNFIHNFAVAVFVYV
 VTLFRKLTLTIATILTLNPLVLYQYTSRTGYTVYVLTILLIIVYTKNNIRKTEVY
 APYIDILFLATPLCSTIEFNSNFVQKDSLLTQRLNIAHILQYVDGILTPNSKESK
 VLEFDSYSLATSMYGVLTNFCMIIYYIYSKRVNVEIQILFIMSVLFIESTPES
 VNMSIWMYFGITPCGVDDLQREFTWANKN"
 7875..7880
 /gene="cpsIVM"
 7949..8800
 /gene="cpsIVM"
 7949..8800
 /gene="cpsIVM"
 /codon_start=1
 /transl_table=11
 /product="putative glycosyltransferase CpsIVM"
 /protein_id="AAK43610.1"
 /db_xref="GI:13876778"

Db	13692	GCAATGGGTGCACAAAGTATTATGAAAAACACTTTACTTTAGATACATAATATGAAAGGTCCG	13735
Qy	721	GATCATTTGGCCCTCAATAGAACCTGATGAACTGAACACATCTTTGATTTGGGCTCAGGTGT	780
Db	13752	GATCATTAAGCCAGTGTGCACACCTGATATTTTAGCTGCTTTAGTAAAGGGGTTCCCAT	13811
Qy	781	GTTGAAAAATCTTTAGGTCAAAATAGTAAAGTGTTCACAGCTTCAGAGTCAAGAAAGAAATAA	840
Db	13812	GTTGAACAAACCTTAGGTGATTTGAAAAATCCAGATCCAGATCAAGTAAAGAAAAATTAAG	13877
Qy	841	ATGCTGCAAGAAAGCTATTATTTAGCTTAAACAGAGATAAAAAAAGGTGAGTTTTC	900
Db	13872	ATTGTTGCTGTGTAATATAGTGTGCTTTTAAACATTTAAABAGCGATATTTATTC	13933
Qy	901	GAAAAAAATATTAACAACAAAGAACCTGTGATGTATCATGTCCTGATGAGTGTATAT	960
Db	13932	ATGAAAAATATATACGGTGAAGCCGCCAGGTAATGATATTTCTCTATGAACTGGTATGAC	13993
Qy	961	TTATTTGGTAAATATGACAGCAGACCTTATATCCAGATGATTAATTAATTCATAGCGAA	1020
Db	13992	ATCTTGGACACAGAGCGCAAGATGATTTGGAAGAGATGATATTCGTGATTCACGC	1405
Qy	1021	TTCAAAATCAGGGGGGATATATGAGAACAAATATAT	1058
Db	14052	TTTCAAAATCATTTGCCCTAGTTGTTATATTAACACTTAAT	14089
RESULT 4			
AF337958	16448 bp	DNA	linear
LOCUS			
DEFINITION	Streptococcus agalactiae strain NT6 type VI capsular polysaccharide synthesis operon, complete sequence.		
ACCESSION	AF337958		
VERSION	AF337958.1		GI:13022158
KEYWORDS			
SOURCE	Streptococcus agalactiae		
ORGANISM	Streptococcus agalactiae		
	Bacteri: Firmicutes; Lactobacillales; Streptococcaceae;		
	Streptococcus		
REFERENCE	1 (bases 1 to 16448)		
AUTHORS	McKinnon, K., Chaffin, D.O. and Rubens, C.E.		
TITLE	Direct Submission		
JOURNAL	Submitted (19-JAN-2001) Infectious Disease, Immunology and Rheumatology, Children's Hospital and Regional Medical Center, 4800 Sand Point Way NE, Seattle, WA 98105, USA		
FEATURES	Location/Qualifiers		
source	1..16448		
	/organism="Streptococcus agalactiae"		
	/mol_type="genomic DNA"		
	/strain="NT6"		
	/serotype="VI"		
	/db_xref="taxon:1311"		
gene	895..16444		
	/gene_type="VI capsular polysaccharide synthesis operon"		
gene	895..2425		
	/gene="cpsA"		
-35_signal	895..900		
	/gene="cpsA"		
-10_signal	919..924		
	/gene="cpsA"		
RBS	956..961		
	/gene="cpsA"		
CDS	968..2425		
	/gene="cpsA"		
	/note="similar to TycR"		
	/codon_start=1		
	/transl_table=11		
	/product="CpsA"		
	/protein_id="AAK11658.1"		
	/db_xref="GI:13022159"		
	/translation="MSNHSRRHQGHSHTEPLRVNLEFLVIFILISVSLFLMRRHFF		
	LAPHLNVIVGVVITLIIILASLFLCIRKKRIPTTIIIVLSIPVATITVGFSTIDL		


```

TNNLNKTAASYEISEMSVVPKDSKI TNIEAVSKLA PVKNDTSNI TDLI EHKISEKGI
SITPOTDSYDAVYRIKNGDSQAMVYLAELSTPDEKSGIKTIYTYRIKKKI
NKKNTHKXGVNIVYISGIDTSGISTVRSNIVNIEMVNTNTHKVLITTPRDYVVK
IPDGGNOVDKLTNAGLVGERTSMKTEKLYDINDYARINSSFLKIDLGAVYV
VNDQAPTSXZKGNPDPVGVTLNSBOALGVPARYSYOGGNDRGNSOKVIAIINK
LMSGSVTKLNSITTSQLOTSVOTNMITDINDLNNQSTGQRTVESQALTGSGSTG
ELPSYAMPQAQLYMSIDDSLSNAKSKIKNTER"
gene      2415..3162
RBS       /gene="cpsb"
          2415..2421
          /gene="cpsb"
          2431..3162
          /gene="cpsb"
          /codon_start=1
          /transl_table=11
          /product="tyrosine phosphatase Cpsb"
          /protein_id="AKI1659.1"
          /db_xref="GI:13022160"
          /translation="MIDISHIVPVDVDDGPKTLZSLSYIEESVQGVRIIVSTSHR
          KGMFEEDDITFKNSIVYHSEAKRFESHQILYGGELTYTSDMLEKTKQIPTANT
          KALIFSWQTSWDHTALSNVLMGITPVVAHIERVNALENQKEREVEIINMGCYT
          OINSHILKQKLFNDHQKFKRARYPLEENVHPVAPSDMENLDVNPPLAAYATIC
          RDGKERANQLFIENQSLIKNHYI"
          3161..3863
          /gene="cpsc"
          3161..3167
          /gene="cpsc"
          3171..3863
          /gene="cpsc"
          /codon_start=1
          /transl_table=11
          /product="putative chain length regulator Cpsc"
          /protein_id="AKI1660.1"
          /db_xref="GI:13022161"
          /translation="MKRIANTEVEINIFMLEKLMKKKFLITPVAAIAPTAGLPSLE
          IVTPQYTSSTRIVINPTNPNNSITQDIOQSGFLANDYKEIITSTVLEKVISSEKL
          NYPSSQLQKIVISILKDTKIVISIVEDANPKMSQKLANSVREAAVSKAVTQVEDI
          TLEKGNLPKAPSSPMKKXNVLIGFIVAGLSITVIVINGILDDRVNTBEDIEKVLGL
          TSGIVPDLNKL"
          3863..4563
          /gene="cpsd"
          3863..3868
          /gene="cpsd"
          3874..4563
          /gene="cpsd"
          /codon_start=1
          /transl_table=11
          /product="capsular polysaccharide chain length
          regulator/exporter Cpsd"
          /protein_id="AKI1661.1"
          /db_xref="GI:13022162"
          /translation="MTRLRIVDSKLRQAKKTEEPYFPAIRTNIOFSGENKILATSVR
          EHGKSTSTSLALSIAQGFKLIDACTRSVMSGTKRATGILKGLINVSQANDL
          GDICETNPRLMVSPSKYPRPPTLLIQNAYMKKTAIKNIPDYLIIIDTPRIGIVY
          DNAILASACDGFVLVTOAGRIKXNYEKAQMGSGSKFLGIIILAKVMSVATIGDY
          GNYGKDRRK"
          4577..5964
          /gene="cpsb"
          4577..4582
          /gene="cpsb"
          4615..5964
          /gene="cpsb"
          /codon_start=1
          /transl_table=11
          /product="glucose-1-phosphate transferase Cpsb"
          /protein_id="AKI1662.1"
          /db_xref="GI:13022163"
          /translation="MIGVYVYVSVSLTLLITTPNKSNDLFLVLLIHVIVFYLDF
          YRDFMSGYIEFRKMLKISFYITFISSSLFPIPKNSPTTTLSPFTPIANSSILLYL
          LNSFLKRYKYSYAKSRDQKVLITNKSLSMTRFKNDHNYIAVGLDSSSEKQV
          DLHNSLRIRINKDALISBLCLVDDAFINIPILFGKYQIDIIIDIEAMGVIVAN
          VEAISPDNIGEKRIQTFEGSVITYEMKRYKSLAKRFIDMGALIGLLCGIVAI
          FLVPOIRKDGPAIPSONRYGNGRIFFRYFKRSKVDQAIQIKDLVHNMQNTGLMFK

```

```

LADDERITKICKIRKITSIDELPOFYNVLKQDMSLVGRPPTVDEYKNTQKRLS
FKPGYTGMOQSGRRNIDPDEIVKLDVQYINEMSIWBDIKIILLKRVLLGTAK"
gene      5975..6437
RBS       /gene="cpsf"
          5975..5981
          /gene="cpsf"
          5988..6437
          /gene="cpsf"
          /codon_start=1
          /transl_table=11
          /product="beta-1,4-galactosyltransferase enhancer Cpsf"
          /protein_id="AKI1663.1"
          /db_xref="GI:13022164"
          /translation="MKITLVSSSGCHLAHLNLRPKMEKDRFWYTPKEDARSLIRE
          EIVHCFPPTRNKNLVKNITLAKRVLRKRPVITISSGAANVPFYIGLFGCKT
          VIVEVFRIDKPYTLGKLVYFVTDKFIYQWEEKKIYKALNLGSLP"
          6422..6910
          /gene="cpsg"
          6422..6428
          /gene="cpsg"
          6437..6910
          /gene="cpsg"
          /codon_start=1
          /transl_table=11
          /product="beta-1,4-galactosyltransferase Cpsg"
          /protein_id="AKI1664.1"
          /db_xref="GI:13022165"
          /translation="MIFVTYTHQEQPRLIKENDRLKGTGAIDQEVPIQNGYSPFEP
          QNCQSKRLSTDQMSWKKEASIVITHEGPATFMSVLSGLPYVPPRKQSGEIND
          HOIQPLNSIAHLVPLAMIEDVDGLAALKRNIAEKYQGNNDMPCHLKKIIGEL"
          6898..8064
          /gene="cpsH"
          6898..6903
          /gene="cpsH"
          6907..8064
          /gene="cpsH"
          /codon_start=1
          /transl_table=11
          /product="capsular polysaccharide repeating-unit
          polymerase CpsH"
          /protein_id="AKI1665.1"
          /db_xref="GI:13022166"
          /translation="WRKTLDDYSLPYALMLIVPNQYQFLITTIIVLLMKSEF
          RISLNSNLSILFLMLFYLFAILIRGVEDITPQFIABLLKLTSTGALPFTNYRK
          ADPNSSVVRNVAKNVYEVLPITVYALFEPNEFTFLCRDLSIMISMKRVLAIVP
          EYATLLGQFIFETVYILFLKQORYGEMIFITLFFVCALYGARIFLQMLILGLYL
          LEIINKFNKLTCKAVPLIIGIILLVCFSSKYSILNYIHYRFQSSSTPLTVY
          ESIRAILGNFLIGQIRVPSSVGIPLGSHSYSISIFRTSPTGLFFESILLFYRE
          AIKQRIITYKFLPGLILLMYFBERDPNHSVLLFTLLGIVGRNDKKS"
          8034..9013
          /gene="cpsVII"
          8034..8040
          /gene="cpsVII"
          8045..9013
          /gene="cpsVII"
          /codon_start=1
          /transl_table=11
          /product="beta-1,3-glucosyltransferase CpsVII"
          /protein_id="AKI1666.1"
          /db_xref="GI:13022167"
          /translation="MIKLVAVIVAVNSBLVENCVESLLQOYPEIEIILLIDGST
          DKSSHICNPLKPDGRVYVAKVNGASASANVGLIEMAGEEITIVDSGDVVALMIE
          IMANLLITENADISIDPEVSDDPYRKQRKRGYVPRGNKSLKEEPKQKENVVWG
          KLYKSLIGDLFNSKRYKIGEDLNFQILNKREHYIVDTRSLIYTIIEKSTANQO
          FNKTLIDIDIFNBIHQOSPELEFVYAKVREKIKLRRKFSGLFADENLRLORI
          KFMQDIKYSYICKAIRFLSKKHICTLYLMKFPYVYIKYNNFKQO"
          9031..9752
          /gene="cpsVIJ"
          9031..9037
          /gene="cpsVIJ"
          9060..9752
          /gene="cpsVIJ"
          /codon_start=1

```



```

source
1. .17276
/organism="Streptococcus agalactiae"
/mol_type="genomic DNA"
/strain="COH1"
/serotype="III"
/db_xref="taxon:1311"
/complement(7. .930)
/gene="cpsy"
/complement(7. .930)
CDS
/gene="cpsy"
/feature="LysR homolog: putative transcriptional regulator"
/codon_start=1
/transl_table=11
/product="Cpsy"
/protein_id="A053064.1"
/db_xref="GI:5823210"
/translation="MRIOQVYVIRKIVSGMNEAAKQVITTPSLSNVRNLETEMG
IOIPIRPRKIGITLTKGMEFLSVAROIIEPTALREYKGDNTSRELFSQHVAV
VNAFVALPNTGDMTOYELPIRETRTWIITIDVQPSSEIGVLEISNVDVITKLPDD
NSLIATLTFTTTHIFVSKNSPLANKSKQLEDPYISTQGLANSFYREEMMS
QIHPKSIIVSDRATLFLNMGIDGVTAGILNSKLNGEIVAIPIPDVDIVIT
RHDKANLSKMGQRFIDYLEEVSFNHKKQS"
932. .951
1047. .1052
1070. .1075
1083
/feature="repeat region
-35_signal
-10_signal
misc_feature"
/feature="CpsIII operon transcriptional start site"
/evidence=experimental
1108. .2580
/gene="cpsa"
1108. .1114
/gene="cpsa"
1120. .2580
/gene="cpsa"
/feature="LysR homolog"
/codon_start=1
/transl_table=11
/product="Cpsa"
/protein_id="A053065.1"
/db_xref="GI:5823211"
/translation="VMSNHSRRHQKSHSTPLRVINLFLVITFILSVSLFLMYRHF
LAPHLNVIYGVVILILILASLFLCIKNAKRIPTTITLVLSIFVATTLTGKSTIDL
TNMKNATASYSLEMSVVPDSDKTNIEKAVSKLAAPVNDTSNITDLIRIKSEKGI
SITPOKDSYODANRIRKNGDSQAMVNAVSILISTDPFSQIETITIVYIKKTI
NRKTNKEGVNINISGIDTFGSIISTVSRSVDIINTVNTKHYLLTTPDAVYK
IPGGGNOYDKTAGLGVGETSKTLENLYNLDVYARINFSPEKIKRPGRSIS
LYTKLIQVKNVITIDFPVQVTLNSEQALGFRKRSYSLGGDNRGNGKRYAAIIN
KLASSOVATKLNSTISQLOTSVOTNMTINDLINQISTGQRPVESQALVGHST
GELPSVMPGQLYWMSIDQSSLSNAKSKIKTWEE"
2570. .3317
/gene="cpsb"
2570. .2576
/gene="cpsb"
2586. .3317
/gene="cpsb"
/function="tyrosine phosphatase"
/codon_start=1
/transl_table=11
/product="Cpsb"
/protein_id="A053061.1"
/db_xref="GI:153590"
/translation="WIDIHSHIVPDDDEPTLEBSLSLEBSYRGVRLIVSTSHR
KGMETPEDIIKFNPSIVKHETERFHLQLLYGGELYTSMLKSKLKOIPLANT
KFLIEPMSQTSMDIHTALSNVIMLGITPVVATIERVNALENQKREKVKIINNGCT
QINPHILKQKLFNDKKEKFKKARAYLEBVLVHFVASDMENLDVRPPFLAEVAKIIC
RDPGERANOLFENAGSILKNHYI"
3316. .4018
/gene="cpsc"
3316. .3321
/gene="cpsc"
3326. .4018
/gene="cpsc"
/feature="putative polysaccharide chain length regulator"

```

```

/gene="cpsd"
4018. .4718
/gene="cpsd"
4018. .4024
/gene="cpsd"
4023. .4718
/gene="cpsd"
/function="autophosphorylating tyrosine kinase"
/feature="Cps chain length regulator/exporter"
/codon_start=1
/transl_table=11
/product="Cpsd"
/protein_id="A053063.1"
/db_xref="GI:1323730"
/translation="MTRLEIVDSKURQAKTEETNAIRTNTPQSGKKNKILIASVR
EBSGKSTISLALSIAQGFYLLIDADTRGVSAGTKAGTIGKLTNYLSGNADL
GILICEINVRIMVVPKGVPPNPTALQNAFNNKIMIAIKNIIDYIITDPIGLVAV
DAIITANAGCGIIVTQACRIKRNVEKXKQMEQSGSKFLGIIINKNVESVATYGDY
GNVGRKRRKX"
4755. .6119
/gene="cpsf"
4755. .4761
/gene="cpsf"
4770. .6119
/gene="cpsf"
/function="glucosyl-1-phosphate transferase"
/feature="Glucosyl transferase, previously described as Cpsd"
/citation=[1]
/codon_start=1
/transl_table=11
/product="Cpsf"
/protein_id="A053066.1"
/db_xref="GI:5823212"
/translation="MIOVYVPSASITLITLTPNRSXNDLFLVILHYIVYISDF
YSDPWRGRLGSEPRKWLKTSFYIFISLFTFKNSFTTTLSTFTPIANMSIILYL
LNSFLWPKRGVYAKSRDQKVALITNKSLSKMPFRNRYDHYNAICLDSKQCT
DKHNSLARIINDALISELTCVVOQAFINIPILFGKXQIODIINDISAKGVYAVN
VEALSPTNIGERKIOFEGSVITYSMKRYKSHIARAFIDITGAIIIGLIGIVAI
FLYPOIRKGGPAIISQNRVGNRGRIFFYKERSHVDLQIKDILVKNOMGTLMK
LDDPRITITGKGFIRTSIDELPOFNVTKGMSLAGTSPVDEYKNSYOKRSL
FRGITGLWQISGRNNTTIDFETVLDVQYINEMSIWSDIKIILLTLKVVLDGTAK"
6131. .6592
/gene="cpsg"
6131. .6137
/gene="cpsg"
6143. .6592
/gene="cpsg"
/function="beta-1,4-galactosyltransferase enhancer"
/codon_start=1
/transl_table=11
/product="Cpsg"
/protein_id="A053067.1"
/db_xref="GI:5823213"
/translation="MKICLVGSGGGLAHNLKPIWKEDEDRPWTEDKEDARSILRE
EIVYHFPFPRNRYVRLVYKTIAPVLAKEKRPDVIISGAAYAVFPFYGKFGCCT
VIVLEVDRIDKPLTGLKLVFVYDKETVQMEKKVYPKALINIGLIF"
6577. .7065
/gene="cpsg"
6577. .6583
/gene="cpsg"
6582. .7065
/gene="cpsg"
/function="beta-1,4-galactosyltransferase"

```



```

/product="capsular polysaccharide F"
/protein_id="BA033746.1"
/db_xref="GI:3721916"
/translation="MKICLVSSGGHLLMLNLPKPKGNDRFMTEDKDAASLIRE
EIVYCFPPTRNRYKLVNNTLLAFKTLREREDVLISSGAVAPEFFYIGKLGCKT
VYIEFGRIDKPLTGKLVVPTDKFIVQEEKKKVPKAINIGGIF"
2266..2808
/gene="cpsP"
2266..2808
/gene="cpsP"
2266..2808
/gene="cpsP"
/codon_start=1
/translation="galactosyltransferase"
/protein_id="BA033747.1"
/db_xref="GI:3721917"
/translation="MTFVYGTTHQOQNRILKEVDRLKGTGALIDQVFIOTQSDPEP
QNCQMSKFLSYDDMSYKAEIVITHGPRATWNAVSKKKTIVPRQOPGEHVN
HODFLKELFKELDYILNISELNIIXEKNIISTKSVISQNDPCFSFNEHFVYL
NKYLLEKKEIENISQISIC"
2847..3986
/gene="cpsB"
2847..3986
/gene="cpsB"
/codon_start=1
/translation="CP polymerase"
/protein_id="BA033748.1"
/db_xref="GI:3721918"
/translation="WQILLALIVLLICSSYKKMKFLYMAIFPIVPIYITSL
LHSLFKTPDPRIILAFNSLITIGVVALKRWYNTLLELQKILKALFNGLLFELG
GLYVYCHMNIQNIISIFGRDLISDMNGHGTQAFMGFEYSNLIIPMTVNTIYIY
YMKLRNYSIMTIGVLLFTFLPIYSGSAGIYVAILQMPILMLNTVVAKKIKFLI
YILFLLVIMVLLFNDLSIYRIINLRSGSSSRSVYKDYIVINNSLLFGESV
KELNLSPLGSHSTIYGYKSGGLIMNIVGYLLFINIGKKAQSAFYELIYG
TLITLSPFALEDIDGANMLIVIFLTVLGIENDYFSLKRWKS"
3986..4987
/gene="cpsH"
3986..4987
/gene="cpsH"
/codon_start=1
/translation="N-acetylglucosaminyltransferase"
/protein_id="BA033749.1"
/db_xref="GI:3721919"
/translation="MEKRIIVSIIPIYNSAYLKECVQSYLQGTPLIEVILIDGS
TMSGECIDNLSQEDNILLVFKHKGVSANRMLDKSGEFTFVDSDPVAPNMI
EIMLKNLITENADIAVDFOISNRDYKRRKRFYVFNKNSLKEFLSPVANSIV
CTLYKKSIIQNLRFDENLKTGEBLLFNCLCOEHRIVDTTSLYTRIIVKISAKO
OKFENSLDPTTIFNEVSLVPAKLAVYERKRLRRIKICLRKMFELGNSINDKIKVQ
REIFPKDIXSPFKAVKYLISLKLISFTLMKCPKCYWAAVRFKIVABEIGENL"
4992..5939
/gene="cpsI"
4992..5939
/gene="cpsI"
/codon_start=1
/translation="galactosyltransferase"
/protein_id="BA033750.1"
/db_xref="GI:3721920"
/translation="MKIKIEKDLISVIVPIYVEDVYVECIESIIVQTYNIELILN
DGSITNCATILAKESERDQVILIKSNGGSRANVGIHISKGVILFVDSQKSS
DYANLVNLIKHSISAIIGLYEFERHNSIRNYELDKYIPEEALMLMDKYG
SITFAMKLPKHSIFNDLEPALKYHEDERFPAKANSIYIDKPLVYHVRIG
SIMNSDNVITARKKLDVLSADRILKILISLRKVSALIKTEIFVYNQFRTKKFLK
OQSVFKEDNTIYRMTGRLLRKYKLVDKLIKIRFP"
6023..6979
/gene="cpsJ"
6023..6979
/gene="cpsJ"
/codon_start=1
/translation="BA033751.1"
/protein_id="BA033751.1"
/db_xref="GI:3721921"

```

```

/translation="MTNFVCHTLYLHLLITIKLKNKENTRIFICDTITDEYTWKILN
DQIRTESKPERAVRQOLONKXIEEEMELVSDLDNHYPERVPTQVYLFEDDILLGYM
VYLGKRYHLITBGNCPQKALGLSSYKRYVYIKYTLFKYVYVGSKYCLSTIEVSLV
GLPHDIRKQKELERKRLPDSLNKQKSLIKRIKTRKPLTTPPSYLLTTPPLADPK
WKTPTRQOSIOEYDFDDIVQSYRLGVNIVKTRHDVVDVDSIKLEVEELPSNP
MEIILMSGRREGOITHSTALDPLTCDKRTITLVDKIDK"
6976..8376
/gene="cpsK"
6976..8376
/gene="cpsK"
/codon_start=1
/translation="capsular polysaccharide"
/protein_id="BA033752.1"
/db_xref="GI:3721922"
/translation="NNTKRLQSGFVYTLGNLVQGLAFITLPIYTRVISTEVQGS
LYVAMNITMLFPIQOTSGLSAAYVKGEEKYSGSAFVGNIMFLLIILIAELFR
SELPAPGSESIFILMVQCSASYVTTRQYFIQQSRLNLIISLANSYSVALS
LPLIFRMSDPLSRVFGAVPTITGIVAFATITHSKSFNKRIFRIVTAVPLIF
HLHGQLQQLDRMLARLVNTRKAVMSRGSLMI IQIVANSIMAMI PWEDARK
EKILQSTYSRYLYGVFLITGYLVPEELAQIMGDRYSSSVQFISLIIVSYELVF
LYTFPNVIGFVYANTWIDPILGTLAAGVAMLNLVLPYHAAVGAAMATIIISYLLAI
FHHIVSKRYHVSQVSRQYIILSGIVSYAMLMNMFNDNIVIRNELIILIVSIV
EQVYILDLISKKRRR"
8376..9401
/gene="neuB"
8376..9401
/gene="neuB"
/function="sialic acid synthesis"
/codon_start=1
/translation="sialic acid synthesis"
/protein_id="BA033753.1"
/db_xref="GI:3721923"
/translation="MYVILAEICGNHNDINLAKKXVDVAVSGVDVAVRQFRAKEL
ISKPAKAYOKETTGADSOLEMTRLSESEYEMEDVYAIKSGVEFTSPFDES
LBRILSTMDPIYIKPSGRIITNLPYLERIKQKQKVLISGMAVEIEHVAVNLIRNG
TMDISLHCCTEYTPRPBSINLVHTLKDPEFDLIGSDHSIGSEVPAALAMGAE
VIRKFTLLDNMGSPDHKSNTPDILALVYKRIYVQALGRREKTPDPEERKNTVA
RNSVALKPKIKGDIVISINIVKRPNGISIPANWYDILGQEAQDDPEDEVDIRDSRP
ENQLEPL"
9478..10206
/gene="neuC"
9478..10206
/gene="neuC"
/function="sialic acid synthesis"
/codon_start=1
/translation="sialic acid synthesis"
/protein_id="BA033754.1"
/db_xref="GI:3721924"
/translation="MKKICLVGSRABYGIKPELQRLSKQSEVNTQIATAMHEEK
YGYTVAQIBEDGFDIAVKVPLHXYDTRDTVMSRAHLOGLKTRKIDKBYDVLITIG
DRYEMLPVNVNALLIVPCHLHGSETSLANPEBYRHHVTKSHLIVSTEDFORV
IQNGRQPVVINTGALGVENALSIPELTKEALIEKQIGYIABSEFPVYLHPVTFEQQ
SAGEQKAVIALSKFVQCLFIS"

```

ORIGIN

```

Query Match      30.7%; Score 324.8; DB 1; Length 10206;
Best Local Similarity 58.2%; Pred. No. 2,1e-42;
Matches 616; Conservative 0; Mismatches 427; Indels 15; Gaps 2;

QY 1 ATGAGTAATATATATATCGTGTGCAAAATGTTGCAACCAATATGTTAGTATAT 60
Db 8370 AAGATAATGTTATATATATATGACAGATGTTGCAATATATATGAGATATTAACCT 8429
QY 61 GCAAGGAATATATATTAAGCAAGGAGCGGTGTATATGACGATTAATTCACAA 120
Db 8430 GCGAATAAATGTTAGATGTCGCGGTCTCTGCGGTCTTATCTGTTAATTTGACACT 8489
QY 121 TTAAAGCTGATTAATTAATTTAGCTATTGCACTTAAGGCAAGATATCAATTAATAAAC 180
Db 8490 TTAAAGCTGAAATCTAATTTCTAATTTGCTCCAAAGCTGAATATCAAAAAGAAACT 8549
QY 181 AAGGAGAAATTAATCTGCTTGAAGATGCAAAAAGCTTAATTAATGAAGTATGACGAT 240

```


		/gene="cpsvH"	
RBS		6866..6871	
CDS		/gene="cpsvH"	
		6878..7972	
		/gene="cpsvH"	
		/note="putative repeating unit polymerase"	
		/codon_start=1	
		/transl_table=11	
		/product="CpsvH"	
		/protein_id="AAK29654.1"	
		/db_xref="GI:13549132"	
gene		/translation="MKINKNSUPFYAIPLVNPFKSLGIGBENSAXITWLVAILLGGI KFLDLSFPERKAVTILFLFTATILNLPFAHKATPILTLFLPALKDLSIKAKPSII IGSRILGLINQDIPEKLDBLEIKYVNEFRDQFLIRSDLGHPNFTHNFPALTFLX VLANKRRLKPVMVFVLTILNTALYLXYTESRSGYIVLYLIVLYTKSLIKRVEMKI AAYGVFLIFVLFPLISTFEFNPNFOKIDVLIITGLHAHLQLVDGLTPFGNSRETS VLFDNSYSMLSMGVLTMFOMIILYYISKKIILEQLLLPLMSIILFFESYPVS VNNIIMVPRKIFCDGIEPIKEPTIVANI"	
		8001..8736	
		/gene="cpsvM"	
RBS		8001..8006	
		/gene="cpsvM"	
		8011..8736	
CDS		/gene="cpsvM"	
		/note="putative glycosyltransferase"	
		/codon_start=1	
		/transl_table=11	
		/product="CpsvM"	
		/protein_id="AAK29655.1"	
		/db_xref="GI:13549133"	
		/translation="MIPVIHYCMFGNPLDPNLKYLIKTRBOCPDYELIEMENNY DYSKAVPRBATKCRPAVSPYADBLDIITYGGYILDVDVELLSLPLRIHECFIA KEISCVTGLIICGVKHHPFKSNMSITDSDGLSLNKTCVEYTNLLINRGCKKN LIIDDDITPYRANFNPKNLLTGKVDCULTVTVSIHRHBGMSSSTISDLKVRV LIIDFGYGYTRMLRFKLKK"	
		8754..9649	
gene		/gene="cpsvN"	
		8754..8759	
RBS		/gene="cpsvN"	
		8762..9649	
CDS		/gene="cpsvN"	
		/note="putative glycosyltransferase"	
		/codon_start=1	
Query Match	30.7%; Score 324.8; DB 1; Length 18239;		
Best Local Similarity	58.2%; Pred. No. 1.8e-42;		
Matches 616; Conservative	0; Mismatches 427; Indels 15; Gaps 2;		
Dn	1 ATGAGTAATAATATATCGTTGGTGAATTTGTGCACCAATATGTAGTGTGATATT 60		
Gy	61 GCAGAACAATGATTTAAAAAGCCAAAGGCCGGCTTAATAGCAGTAATAATCCAACA 120		
Dn	13938 AAGATATGTTTTATTTATTTGACAGCATTTGCTGCACATATATGAGATATTAATCTT 13997		
Gy	121 TTTAAAGCTGATAATTTAATTTTCAGCTATTCGACTTACCTTAGCGAGATCAATTAAMAAC 180		
Dn	14058 TTTAAAGCTGAGAACTTATTTCTTAATTTGCTCCCAAAGCGTAATATCAAAAAGCAACT 14117		
Gy	181 ACAGAGAAATTAGAAATCTCAGTTAGAAATGACAAAAGCTGGAATGAGATGACGAT 240		
Dn	14118 ACAGAGAACGACACAGCTCAACTTGGAGTAGAGAGAAACGTTTAGAGTTAAGCTTAGAGAA 14177		
Gy	241 TATCTCATCTATATGGAATATGACGACGCTTAATTAATTTAGANGTTTCTTACCCTCTT 300		
Dn	14178 TACTTAGAATGCGTGATTTATGCAATTTCAAAAGTGTTGAGACCTTTCAACACCTTTT 14237		
Gy	301 GACCAAGACTCTATATATTTTATAGCACTTTGAAACAAAATAATAGAAATCCCTTCA 360		
Dn	14238 GATGAAAGACTCATTAAGCTTTCTAATTTCTTACAGATATGCCAATTTTACAAAATTCATCA 14299		
Gy	361 CGTAGATATGAATTTACCGATCTCTGAAAAAATATGCAAGCTTCCGATCCCTGATAAG 420		

Db 14298 GGAAGATCACTAATTTACCTTACTTGAAGAGTTGGCAGAGAAC-----AAAAG 14348

Qy 421 AAAATATCATATCAACAGAGAAAGGCTACTATTGATGAGATTAACAGCTGTCTTAT 480

Db 14349 AAGGTTATCTTCTTGAGAGGGTATGGCGGTATGGAAGATCATCAAGCGGTGATAT 14408

Qy 481 TTTATATAATATAAAGTTCCGGTGTGTAATATTCATATATACCTTACTGATAT 540

Db 14409 TTACGTGAGATGTATACACCG-----ACATTTCTATTTTACCTGTGTACACAGAGTAC 14462

Qy 541 CCAAGCGCTTGTGAGATGATTAACCTTATGCTATTAATGATTTGAAGAAACCTTCCCT 600

Db 14463 CCAAGCCTTACCCCTCTTAAATTTAAAGCTTATTCATCTTTGAAGATGATTTAA 14522

Qy 601 AAGATATACATAGGCTTCTGTATCATTTTACGCGGTTTATGCACTATTTGCGCGCTG 660

Db 14523 GATTTAAGATAGGTTATTCGATCATTTCAATTGATGATGAGATGATGATGATGATGAT 14582

Qy 661 CCTATATGATTAATCTTTATTTGAAAAACATTTCACTTATGATTAATCTATGTCGCGCA 720

Db 14583 GCAATGCGTGCAGAGATTTATGAAAAACCTTTTCTTATGATTAATGAAAAAGTCCG 14642

Qy 721 GATCATTTGGCTCATATGAACTGATGATGATGATGATGATGATGATGATGATGATGAT 780

Db 14643 GATCATTAAGGCTGATGCAACCTGATATTTTACCTGCTTATGATTAAGGCGTTCGAT 14702

Qy 781 GTTGAATAATCTTTAGTTCATTAATGATGATGATGATGATGATGATGATGATGATGAT 840

Db 14703 GTTGAACAGGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 14762

Qy 841 ATCCGATGACAGAAAGTCTATATATGCTTAAACAGAGATTAATAAAGGATGATTTTCA 900

Db 14763 ATTTGCTGCTATATGCTGCTGCTTAAACCAATTAATAAAGGATATTTATCA 14822

Qy 901 GAAAAAATATATACACAAAAAGACCTGCTATGATGATGATGATGATGATGATGATGAT 960

Db 14823 ATAGAAATATTTAGCGTATGATGATGATGATGATGATGATGATGATGATGATGAT 14882

Qy 961 TTAATGGTAAATTTGACGACCAAGCTTATTCAGATGATTAATTAATCATATGACGA 1020

Db 14883 ATCTTGGACAGAGAGCCAGATGATTTCAAGAGATGAGATGATGATGATGATGATGAT 14942

Qy 1021 TTTCAAAATCAAGGCGGATTAATGAGACAAAATTAAT 1058

Db 14943 TTTGAAATCAATTCGCCGAGTATTAATACACCTAAT 14980

RESULT 8
AE014245/c 21365 bp DNA linear ECT 23-SEP-2002

LOCUS Streptococcus agalactiae 2603V/R section 55 of 100 of the complete genome.
DEFINITION AE014245 AB009948
ACCESSION AE014245.1 GI:22534189
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
1 (bases 1 to 21365)
Tetzelin, H., Masignani, V., Cieslewicz, M. J., Eisen, J. A., Peterson, S., Wessels, M. R., Paulsen, I. T., Nelson, K. E., Margarit, I., Read, T. D., Madoff, L. C., Wolf, A. M., Beaman, M. J., Brinkac, L. M., Daugherty, S. C., DeBoy, R. T., Durkin, S., Kolonay, J. F., Umeyam, L. A., Madupu, R., Lewis, M. R., Radune, D., Fedorova, N. B., Scanlan, D., Khouiri, H., Mulligan, S., Carthy, H. A., Cline, R. T., Gill, J. J., Scarselli, M., Mora, M., Iacobini, E. T., Brestoni, C., Galli, G., Telford, J. L., Vegni, F., Malone, D., Rinaldo, D., Rappunli, R., Mariani, M., Kasper, D. B., Grandi, G. and Fraser, C. M.
Complete genome sequence and comparative genomic analysis of an emerging human pathogen, serotype V *Streptococcus agalactiae* Proc. Natl. Acad. Sci. U.S.A. 99 (19), 12391-12396 (2002)

PURGED 12200547
2 (bases 1 to 21365)
REFERENCE
AUTHORS
Tetzelin, H., Masignani, V., Cieslewicz, M. J., Eisen, J. A., Peterson, S., Wessels, M. R., Paulsen, I. T., Nelson, K. E., Margarit, I., Read, T. D., Madoff, L. C., Wolf, A. M., Beaman, M. J., Brinkac, L. M., Daugherty, S. C., DeBoy, R. T., Durkin, S., Kolonay, J. F., Umeyam, L. A., Madupu, R., Lewis, M. R., Radune, D., Fedorova, N. B., Scanlan, D., Khouiri, H., Mulligan, S., Carthy, H. A., Cline, R. T., Gill, J. J., Scarselli, M., Mora, M., Iacobini, E. T., Brestoni, C., Galli, G., Telford, J. L., Vegni, F., Malone, D., Rinaldo, D., Rappunli, R., Mariani, M., Vegni, F., Malone, D., Rinaldo, D., Rappunli, R., Telford, J. L., Kasper, D. B., Grandi, G. and Fraser, C. M.
Direct Submission
Submitted (18-JUN-2002) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
Location/Qualifiers
1. 21365
/organism="Streptococcus agalactiae 2603V/R"
/mol_type="genomic DNA"
/strain="2603V/R"
/serotype="v"
/db_xref="taxon:208435"
/complement(69..1094)
/gene="neub"
/note="synonym: SAG1161"
/complement(69..1094)
/gene="neub"
/note="identified by match to Pfam protein family HMN PF01354"
/codon_start=1
/transl_table=11
/product="N-acetyl neuraminic acid synthetase Neub"
/protein_id="AA00043.1"
/db_xref="GI:22534190"
/translation="MWIIIAEIGCHNNDINAKMVDVAVSGCDVAFQCFPEKTLISKFAKAYOKATVTGSDSLEMTKELSLSEYELKMDVATSKSEVFSTPDEBSLEFLISTDMPYIKPGEITNLPLYEKIKQOKVLTSTGMVMSIHQAVNLKONGTTDLSITHCETPEPPEPSLNTLVHTLDEDFKLTIGYSDHSIGSEIVAAAMKAEVIEKHETLDNMGEPHRSATPDILALVKGIVYQALGPEKIPDEVEENKIVA KESVAALXPKIKDITISINITYKRGNISPMWVDILGSEADDFEEDVLRDSRF ENQLEPL"
/complement(1094..2494)
/gene="cpsL"
/note="synonym: SAG1162"
/complement(1094..2494)
/gene="cpsL"
/note="identified by match to Pfam protein family HMN PF01943"
/codon_start=1
/transl_table=11
/product="polysaccharide biosynthesis protein CpsL"
/protein_id="AA00044.1"
/db_xref="GI:22534191"
/translation="MNTKILSGCFIYTLGNILVQGLAFITLPIYTRVISTEYGGYS LTVAMGILTLPIGLQTSGLSASARKYGRGHEPESYSGASPSVNIHFILVLAIFER SFLAPLVGFSSEIFLAVCOQSTANVYVTFEGQFTQOQRELANILISLNAVSVALS LFLIFMSDDPLSRVFGAVPTITIGIVAFATYVYSKSPNKRIFVIVSLEPLF HLGHQILNQLDRIMLARLYSTREVMVSGVSLNMIQIVLMSINMAIMPEFDDAK BKLLQSTVYSRYLYGLVFLVGLVLPFLQIMGDKYSSVQPSLIIIVYFLV LMPYVNIQEVANTWPIPIGLTAGVWMLNLVLIPIYAAAGAAATISYLAII FHHIVKVKHYSDVSGVROIIISGLVFSYMLMNNPFGNIVIRMSIITLIVYSIV FQVYIIDLKRRRRR"
/complement(2491..3447)
/gene="cpsK"
/note="synonym: SAG1163"
/complement(2491..3447)
/gene="cpsK"
/note="similar to GP:13549137, identified by sequence similarity, putative"
/codon_start=1
/transl_table=11
/product="polysaccharide biosynthesis protein CpsK(v)"
/protein_id="AA00045.1"
/db_xref="GI:22534192"

Db	575	CCAAACCCCTTACCCCTCTCTCTAAATTTAAAGCTTATTCATCTTGAAGAGATTTAA	516
Qy	601	AAGATATACATAGGCTTCTGTATCATTTCPAGCGGGTTTATGACAGCTATTGCGCGGTG	660
Db	515	GATTTAAAGATAGGTTATTCGATCATTTCAATTGGATCAGAAGTACCTATCGACAGCA	456
Qy	661	CCTATGCAATACCTTTTATGAAAACATTTCACTTTGATTAATCTATGCTGGCCCA	720
Db	455	GCAATGGGTGAGAGATTTGAAAACATTTACTTTAGTATCTAAATATGAAAGGTCCG	396
Qy	721	GATCATTTGGGCTCAATGAAACCTGATGACCTGAAACATCTTGTATTTGGGCTCAGTGT	780
Db	395	GATCATTAAGCCAGTGCACACCTGTATTTTAGTGTCTTTAGTTAAAGGGGTTCCAT	336
Qy	781	GTTGAAAATCTTNGGTTCAATATGTAAGTGGTTACAGCTTCAGAAAAGATATAA	840
Db	335	GTTGAAACAAGCTGTAGTATTTGAAAATTTCCAGATCCAGTGAAGAAAATATAG	276
Qy	841	ATCGTAGAAGAAATCTCTATATAGCTTAAACAGAGATTAATAAAGGATGAGTTTTTCA	900
Db	275	ATGTGTGTCGTAATCAGTCGTGCTTTAAACCAATTAATAAGGAGATTTATTTCA	216
Qy	901	GAAAAAATATPACACAAAAAAGACCTGTATGTATCAGTCCAGTCCAGATGTATAT	960
Db	215	ATGAGAAATATTTACGTTGAGGCCCAAGTATGTATTTCTCCATGACGTGATGAC	156
Qy	961	TTATTTGGTAAATATTCAGAGCAGACCTTTATTCAGATGAATTAATATTCATAGCGAA	1024
Db	155	ATCTTGGAGACAAGAGCCCAAGATATTTTCGAAAGAGATGAAGTATTTCTGTATTCAGGC	96
Qy	1021	TTCAAAATTCAGGGGAGATTAATGACAGCAAAAATATTT	1058
Db	95	TTTGAAATTCATTTGCCGAGTTATTAATPACCTTAT	58

```

gene
1. .375
/gene="cpsIas"
<1. .375
/gene="cpsIas"
/function="putative histidine protein kinase"
/codon_start=1
/transl_table=1
/product="CpsIas"
/protein_id="BAB62877.1"
/db_xref="GI:15186710"
/translation="VDHYDMFKELKSNRPVLARSERWHDRLTSKSLAECLQTLTD
LDSEVAKSTIIQLRPFKSRVNPALHLIKPDEDELRYTTFQSSQAFIDQLIFIAKTVIG
VEYDTPNHYDTVKRLIKILE"
/complement(400. .1323)
/gene="cpsIar"
/complement(400. .1323)
/gene="cpsIar"
/function="LysR-type transcriptional regulator"
/codon_start=1
/transl_table=1
/product="CpsIar"
/protein_id="BAB62878.1"
/db_xref="GI:15186711"
/translation="MRLIQQLQVIRKIVTGSMTNFAQLYITQPSLSNAVRNLETGK
IQIFLRNGKGTLLTKDGEFLFSVARQLBQVALLERKYGNTSRLEFSVSSQYAPVA
VNAFPAALRNKQDQYELFLBETRTWEIIDVKNFSRIGVLPLANSYRDLATLIPDD
NSIATILFTTTPHIFVSKSPILANKKGLSMKQLEDYPSLSDGLNLSFFSEEMKSS
QIRHFKSIVSDRATLEMLMGDGYTAAGILNSKLNGBRITVALPDLVDSDIIVYIY
RHDKNLSMKQQLIDVLEEVSPRNKKKS"
/complement(1447. .1452)
/evidence=experimental
1453. .1457
/evidence=experimental
/complement(1470. .1475)
/evidence=experimental

```

```

RESULT 9
AB028896
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source

```

-10_signal
-35_signal
gene
CDS
gene
CDS

/element=experimental
complement (1481..1486)
/evidence=experimental
complement (1507..1512)
/evidence=experimental
1512..2969
/gene="cps1A"
1512..2969
/gene="cps1A"
/function="regulator"
/codon_start=1
/transl_table=11
/product="Cps1A"
/protein_id="BA82275.1"
/db_xref="GI:5381178"
/tranlation="YSNHSRRQKKHSHTEPLRVNLFVLVFIILSVSLFLMRHRH
LAFRLHNYIGVAVILILILASHFLCIRKRAIPTTILVLASIYVATLVYKSTID
TNLTKRAISYSEISVIVPDSKITNTEAVSKLAAPVMDPTSNITDILERIKEBG
STTPKTSYODAYNRIRKNGSQAMVLNNAVSLIESTPFRSQIKTIYTKIKKKK
NRKNHBBGVYNIYISGIDTGESISTVSRDANIIMPVNTNHKVLITITPRDVA
IDPGNGYDVKITAGIVGTSTKMLTENVLDINDYARINFSSEKILITLDGVTY
YNDQAFYTKRHGRHPDPVYQVTLNSRQALGCPFRERSLOGGNDGRNRNKIATII
LASDSVTKRNTISQDTSVQTNMTIDINIDLINQISTGQRFVESQALTHGST
ELPSYAMPAPQDLYMSIDQSSLSNKKSYIKTWE"
2975..3706
/gene="cps1A"
2975..3706
/gene="cps1A"
/codon_start=1
/transl_table=11
/product="Cps1A"
/protein_id="BA82276.1"
/db_xref="GI:5381179"
/tranlation="MIDHSHVDPVDGPKTLBESLGLIEBSYRQGVRIIVTSRRI
KAPLPEEIIIFKNSIVKHEAKRFEHQLIVGELLYTSMLEKTLKQIPLTANN
KMLTLEPSQSTQMDIHTLNVLMGLTPVVAHERYALENCKRVEKIIINQGYI
QINSHILKQKLFNDKHKRPKKARARYPLEENLVHFVADMTNLIVRPFLAAVYKII

ROFKERANOLFPIENAGSLKRNHYI"
gene 3715..4407
/gene="cpsIac"
3715..4407
/gene="cpsIac"
/function="chain length determination (putative)"
CDS /codon_start=1
/transl_table=11
/product="CpsIac"
/protein_id="BAAB2277.1"
/db_xref="GI:5381180"
/translation="MNKLAITEVEINIPNLKTKKRLITVAIAFAPAGLPSLF
IVPQYISSTRIVINPNTNNISITADQAGFLANDKEIITSTDLKRYSSKL
NYPSSQLKQITVSLKDTREVISI SVEDANPKMSOPLANSVREAVSKIKAVQVEDI
TTLEKXNLPKAPSSPNIKKVVLIGFIVGAGLSTIVLVMGIIIDRVTEDIEKALGL
TSQIVPDLNKL"
misc_feature 4418..14922
/note="corresponds to DBJ Acc# AB017355"
gene 4418..5107
/gene="cpsIad"
4418..5107
/gene="cpsIad"
/function="chain length determination (putative)"
CDS /codon_start=1
/transl_table=11
/product="CpsIad"
/protein_id="BAAB2278.1"
/db_xref="GI:5381181"
/translation="MTRLEIVDSKLRQAKTEEFNAIRTIQSGENKILATSVR
EGHKSSTTSIALSLAQAQFTLLIDADTRNSVMSGTFRAGTIGLITNYSQNDL
GLDICTNVPRLMVSQKVPBPRTALONAYNKYLBALIFDYLIIIDTPGIVY
DAIILMACDGFILVTQAGRIKRYEKKAEQMGSGSKFLGIIILNKVESVATGDI
GNYGKDRKRX"
-35_signal 5243..5248
/evidence=experimental
5243..5248
/evidence=experimental
-10_signal 5265..5270
/evidence=experimental
5265..5270
/evidence=experimental
gene 5334..6509
/gene="cpsIac"
5334..6509
/gene="cpsIac"
/function="glucosyltransferase"
CDS /codon_start=1
/evidence=experimental
/transl_table=11
/product="CpsIac"
/protein_id="BAAB2279.1"
/db_xref="GI:5381182"
/translation="MVLKYSFYIPISSSLFIFKNSFTTTRLSFSPFIANNISILYL
LNSFLKYRKYSVAKSRDQVYLTKDLSLMTFRNKYDHYIAVCLDSSEKCY
DKHNSLRJIKNDALFSELCTLVDOAFINIPHELKQIOIODIINDIFAMGIVAVN
VRLSPDNIGERIGTPEGYSVITVSMKPKYSILAKFLDITGAIIIGLIGIYAI
FHVQIRKQSGPAIFSCNRVGRGRIFFRYKPSMAYDABQIKDLVHNMGTLMFK
LDDPRTIKGIRKRTSIDELPQIVNLGDSLVGTRPVDREKNSYOKRSL
FKGITGLWQISRRNNITDPELVKLDVQYINMSIWSDIKILLTLKVLGLTGAK"
gene 6533..6982
/gene="cpsIac"
6533..6982
/gene="cpsIac"
/codon_start=1
/transl_table=11
/product="CpsIac"
/protein_id="BAAB2280.1"
/db_xref="GI:5381183"
/translation="MKICLVGSGGSHALNLKPGKXKDRFMTYFDKEDARSILRS
EIVYCPRPNRVNLVKTLLAPVLRERPDVLIISGAAYVAPFYIGLFGCKT
VVI3VDRIDIKLITGLKLYPTVDKFIIVQHEKRYPRALNGLIP"
gene 6982..7524
/gene="cpsIac"
6982..7524
/gene="cpsIac"
/function="galactosyltransferase"
CDS /codon_start=1

/transl_table=11
/product="CpsIac"
/protein_id="BAAB2281.1"
/db_xref="GI:5381184"
/translation="MIFTVGHEQGRNLEKVDRLKGTGALDQEVPIQGVSDPEP
QNCQSKTSLSDYDMSYKMSKIVITTHGPPATFMAVSKKCTIVYPRQGEHYAN
HVDPLKSLFLKELIDYILNISLENIIKENIISTSKYISQNNDFCFSPKHEFIVYL
NKYILKXKIEINISIOSIC"
gene 7563..8702
/gene="cpsIac"
7563..8702
/gene="cpsIac"
/function="Cp polymerase (putative)"
CDS /codon_start=1
/transl_table=11
/product="CpsIac"
/protein_id="BAAB2282.1"
/db_xref="GI:5381185"
/translation="MQLILLALIVLLICSSYKKKMKPLVMAIEPIFYIIVLSIL

Query Match 30.7%; Score 324.8; DB 1; Length 25021;
Best Local Similarity 58.2%; Pred. No. 1.7e-42;
Matches 616; Conservative 0; Mismatches 427; Indels 15; Gaps 2;

1 ATGAGTAATATATATATGCTGTAATGGTGCACCAACCAATAGTAGTGTGATTT 60
13086 AAGATTAATGCTTATATATATATGCAAGATGGTGCACCAATATATATATCTT 13145
61 GCAGAGAAATGATATTAAGCCAAAGAGCCGGTGTATATGACGTAATATTCACAGA 120
13146 GCGAAAAAATGATGATGATGCTGCGTGTGCTGTGATGATGCTGTAATATTCAGCT 13205
121 TTTAAAGCTGATTAATATATGCTATGCACTTACAGCAGATATTAATTAATAAAC 180
13206 TTTAAAGCTGAGAAATCTATATCTTAATTTGCTCCAAAGCTGAATATCAAAAGAACT 13265
181 ACAGAGAAATGATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240
13266 ACAGAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 13325
241 TATCTCATCTAATGAAATATGACGATGATGATGATGATGATGATGATGATGATGAT 300
13326 TACTTAGAATGCGATATATCAATTCAAAGGTGATGATGATGATGATGATGATGATGAT 13385
301 GACGAAAGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
13386 GATGAAAGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 13445
361 GGTGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
13446 GGAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 480
421 AAAATATATATATGCAAGATGCTACTATGATGATGATGATGATGATGATGATGATGAT 480
13497 AAGTATATCTTTCACCGGATGATGATGATGATGATGATGATGATGATGATGATGAT 13556
481 TTTATTAATTAATTAATGCTCGGATGATGATGATGATGATGATGATGATGATGATGAT 540
13557 TTACGTGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 13610
541 CCAGAGCCCTTGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
13611 CCAGAGCCCTTACCTCTCTTAATTAACGTTATGATGATGATGATGATGATGATGATGAT 13670
601 AAGATTAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 660
13671 GATTTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 13730
661 CCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720
13731 GCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 13790
721 GATCATTTGGCTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 780

MEDLINE 22242508
 PUBMED 12354221
 REFERENCE 2
 AUTHORS Glaeser P., Rusniok C. and Frangeul L.
 TITLE Direct Submission
 JOURNAL Submitted (31-MAY-2002) Glaeser P., Institut Pasteur, Genomique des
 Microorganismes Pathogenes, 25 rue du Docteur Roux, 75724 Paris
 Cedex 15, FRANCE. E-mail: pglaser@pasteur.fr Phone: +33 (0)1 45 68
 89 96, Fax: +33 (0)1 45 68 87 86
 Location/Qualifiers
 1. 128050
 /organism="Streptococcus agalactiae NEM316"
 /mol_type="genomic DNA"
 /strain="NEM316"
 /serovar="serovar III"
 /db_xref="taxon:211110"
 complement(join(206..703,713..718))
 /gene="gbs1139"
 complement(206..703)
 /gene="gbs1139"
 /codon_start=1
 /transl_table=11
 /product="unknown"
 /protein_id="CAD46798.1"
 /db_xref="GI:23095614"
 /translation="MSKIQQLIKRDKPLRLTLNSPLHRSQTDCCDYSDDTCDDREYQYE
 RLKGLIVLLKQQLKLSQMSRENRINCLTSYFETIAARRASDSGLIAYILEVLSI
 MEDAAVGMANGIINNEFHALLCQMLNEDYDKLEFKDKNFERFQOKRKHGELF
 EAFGL"
 complement(713..718)
 /gene="gbs1139"
 complement(join(717..1469,1477..1483))
 /gene="gbs1140"
 complement(717..1469)
 /gene="gbs1140"
 /codon_start=1
 /transl_table=11
 /product="unknown"
 /protein_id="CAD46799.1"
 /db_xref="GI:23095615"
 /translation="MMDILSELTKGAGSYNSVNNMTAVNKSIAIGVILVSLFEL
 ELISWYRFRNQGEMTKLFLVAVKYPFLAYVLSQALINAILFELNVTYKLAGV
 DLSGNLFEFKAKGAYGRIRLGNFLAYLSIKWVLAFLFELYTKIAP
 VIVAFMGDDTHRIANPLKTRGAILQSNVILILAILNQGFRVDTRAGLDNSWLT
 YVAISYIKCIIFELILFSGCRKAKQLQVS"
 complement(1477..1483)
 /gene="gbs1140"
 complement(1495..2020)
 /gene="gbs1141"
 complement(1495..1510)
 /gene="gbs1141"
 complement(1518..2003)
 /gene="gbs1141"
 /codon_start=1
 /transl_table=11
 /product="Unknown"
 /protein_id="CAD46800.1"
 /db_xref="GI:23095616"
 /db_xref="SPTREMBL:O8E591"
 /translation="MSESFYVAFADPFKRRKRPPTAELEKANVEGKKKSSRGER
 LRTGSLFVIRFRALLILILQTPVYLTLEFFNLIKSAFAVAVICITIKAVPVIILAEI
 PFSQINDLSQAPKLISFPAQDEMTNNLSPIYTRIDITIIICIFSVFALVMTFSKR
 V"
 complement(2015..2020)
 /gene="gbs1141"
 complement(2109..2351)
 /gene="gbs1142"
 complement(2109..2336)
 /gene="gbs1142"
 /codon_start=1
 /transl_table=11

/product="unknown"
 /protein_id="CAD46801.1"
 /db_xref="GI:23095617"
 /db_xref="SPTREMBL:O8E590"
 /translation="METAVRYVTANGLIALIATVGLQWLVTGAFDYRGRKNSPQMMDO
 GFTSMISGALAVIAGVAAIIVAMRAISF"
 complement(2346..2351)
 /gene="gbs1142"
 complement(2611..5422)
 /gene="gbs1143"
 complement(2611..5409)
 /gene="gbs1143"
 /note="putative peptidoglycan linked protein (LpxTG
 motif)"
 /codon_start=1
 /transl_table=11
 /product="unknown"
 /protein_id="CAD46802.1"
 /db_xref="GI:23095618"
 /db_xref="SPTREMBL:O8E589"
 /translation="MNSQETKSGCFPRKSAVGLVCGIALAGATLTATSQVADQVTT
 QATVTVQNAEIVTSQDLKAVDTAKAAVAVTTTAVNRAITTDQADLANQVQ
 VDVYAKQAQNTQALKDATBAKIDAEKKEAEAEVAAKNAQGAEVDAKNAQAAV
 DARNAKQOADDQKAKIDAEKASQSVSCINQNKAKIDAEKNAQADKADATNAQ
 QNDVQTKLANIKSVAVAGVORRKKQVADANQALQDQYQAKALYNQALAKA
 EADKQINNVAEDIKAKQVQDNAEYKSIWPAKTKPDGSEPNNDMDIGVKTIGYK
 LTGRNHHYVANKQSVTAPVDSVTLTYETRYRANQAANAQNTVPRVLTDRGRPIFE
 KAHNENKIPASTLNTLQNLKTBKPPHSSGNEVFKIHDMVHTDGSALVSVN
 NDVAVNVAIPEQTPPEKPTPAQPKKEHVSVPKIKYKAVHVOQIPKATPTKVLIDENQSI
 VPSIETPTPEKPTPAQPKKEHVSVPKIKYKAVHVOQIPKATPTKVLIDENQSI
 NKSVLPNATLDYVAKNSQKQKASAEIAKAFAPVDQPNEMALAEIVYSIKASN
 GDVSSILMEHVLSKOTLDQKQSLIKAGISPVGEVYMTAPKADPQAFVAYVQGL
 DITVNLSPKRRKEPFGQIKNGVANDIDGNGVNTGNVINDITTPVHNDVLDKREKS
 INNGVYKLGDETYTLDEGVYPARNGYLPFEKPEYDHLQTHEDLYLKRYAKAYATL
 KDGVTPKGNLVQTEFTVYKTEGRYALRADFLAQVSSSAGADPTVYKIKR
 GDVNTADPFPVGNKVKTEITVYTPPEKPKVMPQKVPKAPALPSTGEQSVLIVL
 GAALLSLGLGVFKRQ"
 complement(5417..5422)
 /gene="gbs1143"
 complement(5476..6199)
 /gene="gbs1144"
 complement(5476..6186)
 /gene="gbs1144"
 /note="putative peptidoglycan linked protein (LpxTG
 motif)"
 /codon_start=1
 /transl_table=11
 /product="unknown"
 /protein_id="CAD46803.1"
 /db_xref="GI:23095619"
 /db_xref="SPTREMBL:O8E588"
 /translation="MNRKRLTVLSIAVAVLSVTPADEIPSPDPAFSPPLINPT
 VDRGPTPTKSPVPDGPPTDITKSTPVDPGPTPTKSTPVDSIDNSPDSB
 AITNMIGTKRPDQVYKPEPKIDPVSSTYTPVIDVDKNGVNTGNVINDITTPVHNDVLDKREKS
 TQNGVITRDSGSHVASEVGRHNDGVITTIKSSGKTKTLPRTEBAKFSLSIG
 AILISFVAVLPKQRYTLN"
 complement(6194..6199)
 /gene="gbs1144"
 complement(6201..8447)
 /gene="gbs1145"
 complement(6201..8432)
 /gene="gbs1145"
 /note="similar to surface exclusion protein Sec10
 precursor - Enterococcus faecalis plasmid pcef10 (LpxTG
 motif)"
 /codon_start=1
 /transl_table=11
 /product="unknown"
 /protein_id="CAD46804.1"
 /db_xref="GI:23095620"
 /db_xref="SPTREMBL:O8E587"
 /translation="MNCQIKITGLVATLSAVGNVAEDITTPAPVNEPOVSESTAK
 TPQVTEQVNSAKVTADQATSDVNAQKVNRAQOKQAQOKKVNATTTNETOKLV

Query Match	30.7%;	Score 324.8;	DB 1;	Length 128080;
Best Local Similarity	58.2%;	Pred. No. 1.1e-42;		
Matches 616;	Conservative	0;	Mismatches 427;	Indels 15; Gaps 2.

[illegible]

OY	11	TATATATCGTGTCTGCTAAATTTGGTTGCAACCAATAATGATGTTGGTATATTCGACAAGAA	70
Ds	5	TTTATATATATGCGAGAGATTTGGTTGCATTCATTAATGAGATATATTAATCTTCCGAAAAAAA	64
OY	71	TGATATTTAAAGCCCAAAGGCCGGTGTATATGCAGTATAAATTTCCAAACATTTAAAGCTG	130
Ds	65	TGGTAGATGTGGCGGTGCTTGTTGGATGTTGAAGCTGTAAATTTCAAGCTTTTAAAGCTG	124
OY	131	ATAATATATTTCAAGCTATTTGCACCTTAAGGCAAGATATCAATTAATAAAAAACAGAGAGAT	190
Ds	125	AGAAACTATTTCTTAATTTGGTCCCAAGCTGAATATCAAAAAGAACTACAGAGACAG	184
OY	191	TAGAACTTCAGTTAGAAATGACAAAAAAGCTTGAATGAAGTATGACATTTATCCCATC	250
Ds	185	CAGACAGCTCAACTGATGATGAGAAACGTTTAGAGTTAAGCTTTAAGAAATATCTTAGAA	244
OY	251	TAAATGAATATGCACTGACGTTTAAATTTAGATGTTTTTTCTACCCCTTTTGACAAGACT	310
Ds	245	TGGCGATATTAATGCATTTCAAAAGGTGGAGACCTTTTCAACACTTTTGATGAAGAGT	304
OY	311	CTATGATATTTTACACTTTTGAAACAATAATATGAAAAATCCCTTCAGGTGATAT	370
Ds	305	CATTAAGATTTCTTAATTTCTACAGATATGCCAATTTACAAAAATTCATCAGAGAAATCA	364
OY	371	TGAATTTACCGTATCTTGAAAAAATATAGCCAAAGCTTCCGATCCCTATAGAAAAATATCA	430
Ds	365	CTAATTTACCTTACTTAGAAAAAGATTTGGCAAGCAAC-----AAAAGAAAGTATTC	415
OY	431	TATCAACAGGATATGCTACTATTTGATGATATTAACAGCTGTTCTATTTTATATAATA	490
Ds	416	TTTCACGCGGTATGCGGTATGTAAGAGATCATCAAGCGGTGAATTTTACCTCGA	475
OY	491	ATAAAGTTCCGGTGTGATATATATTAACAATTTACATTCAGATATCTGAATATCCAAAGCCT	550
Ds	476	ATGTACAAAC-----TCACATTTCTAATTTTACATTTGTAACAACAGATATCCCAACCTT	529
OY	551	TTGAGAGATGTAACCTTATATGCTATTAATATTTGAAAAAACAATCCCTTAGATATACA	610
Ds	530	ACCCCTCTCAATTTTAAACGTTATTCACATTTTAAAGATGATATTTAAAGTTTAAAGA	589
OY	611	TAGGCTTCTGATCAATCTACCGGGTTTATGACGCTATTTGGCGGCTGCTTATGAA	670
Ds	590	TAGTTATTCGATCATTCATTAATGATCAGAGTACTCTATCCACACAGACATAGGGTG	649
OY	671	TAACTTTATTTGAAAAACATTTCACTTTAGATTAATCTATGTCGGCCAGATCAATTGG	730
Ds	650	CAGAGTATTTGAAAAACCTTTACTTTGAGATCTAATATGAAAGTCCGAGATCATTAAG	709
OY	731	CCTCAATGAAACCTGATGAACTGAAACAACCTTTGATTTGGGGTCCAGTGTGTGAAAAAT	790
Ds	710	CGAGTGCAACCTGATATTTTAGCTGCTTAGTTAAAGGGTTCCGATTTGTGAACAAG	769
OY	791	CTTTAGTTCAAATAGTAAAGGTATACAGCTTCAGAAAGGAAGATTAATATCGTAGCA	850
Ds	770	CCTTAGTAGATTGAAAAAATCCAGATCCAGTGAAGAAAAAATATAGATTTGTGCTC	829
OY	851	GAAAGTCATTAATAGCTAAACACAGATTAATAAAAAAGTGAGGTTTTTTCAGAAAAAATA	910
Ds	830	GTAATATCAGTGGTCTTTTAAACCAATTTAAAAAAGCGAATTTATTCATTAAGAAATA	889
OY	911	TAAACAACAAAAGACCTGTATATGTTATTCAGTCCGATGAGATGATTAATTTATGGGTA	970
Ds	890	TTACCGTAAAGGCCAGGTATGATATTTCTCTATGATACGTGATATCATTTGGGAC	949
OY	971	AAATTCAGAGACACCTTATATCCAGATGAATTAATAATTCATAGCAATTCAAAATC	1030
Ds	950	AAAGAGCCCAAGATATTTGAAAGAGATGAAGATTAATTCGATTCACGCTTTGAAAAATC	1005
OY	1031	A 1031	
Ds	1010	A 1010	

```

RESULT 13
LOCUS      AB050723
DEFINITION Streptococcus agalactiae cpsbD, cpsbF, cpsbG, cpsbH, cpsbI,
ACCESSION  AB050723
VERSION    AB050723.1
KEYWORDS   GI:12697593
ORGANISM   Streptococcus agalactiae
SOURCE     Streptococcus agalactiae
            Bacteria: Firmicutes; Lactobacillales; Streptococcaceae;
            Streptococcus.
REFERENCE  1 (sites)
            Miyake, K., Watanabe, M. and Iijima, S.
            CpsJ of Streptococcus agalactiae type Ib shows
            beta-1,3-galactosyltransferase activity
            Unpublished
            2 (bases 1 to 9987)
            Miyake, K., Watanabe, M. and Iijima, S.
            Direct Submission
            Submitted (01-NOV-2000) Katsuhide Miyake, Nagoya University, Grad.
            Sch. of Engineering, Dept. of Biotechnology; Chikusa-Ku, Furo-cho,
            Nagoya, Aichi 464-8603, Japan
            (E-mail:miyake@proc.nbio.nagoya-u.ac.jp, Tel:81-52-789-4278,
            Fax:81-52-789-3221)
FEATURES
Source
1..9987
    /organism="Streptococcus agalactiae"
    /mol_type="genomic DNA"
    /db_xref="taxon:1311"
    425..430
    442..1791
    /gene="cpsbD"
    442..1791
    /gene="cpsbD"
    /function="glucosyltransferase"
    /codon_start=1
    /evidence=not experimental
    /trans_table=1
    /product="CpsbD"
    /protein_id="BAB21597.1"
    /db_xref="GI:12697594"
    /translaton="MIGTVAVYFSASITLILIPNFKSNQDLFVLLIHIVYFLISDE
    YRPFMSRGVLEEFPMVIAKSPFYIFISSIFFSKSPFTTRLSFFPIANUSILLYI
    LKNSLRIRKRYSAKFSRDLTKVLLITKSDSKMTFENKXDNYLAICILDSSEKDCX
    DLKNSLRIRKNSALSELFCLEVDQAFINIPILFQVQLODIIINDIEMAGVIVANVA
    VELSFPNIGEKRIQTPREGSVITVSMKPYKSLIAKRFIDMGAILIGLIGCVIYVNA
    FLVQIQRKDGRAIFESGNGVRGNRIIRFQKFRMDAKQDKLDLVHNMQMTLMFR
    LDDPFRITKIGKIFKTSIDSLPQVIVLKGDSLVSTRPTVVEYKSNSTQGRRLS
    FKPRITLQWISGNNTIDEDVAKDQVYINWISDIXITLILTLKVALVIGTAK"
    1803..1810
    1815..2264
    1815..2264
    /gene="cpsbF"
    1815..2264
    /gene="cpsbF"
    /codon_start=1
    /evidence=not experimental
    /trans_table=1
    /product="CpsbF"
    /protein_id="BAB21598.1"
    /db_xref="GI:12697595"
    /translaton="MKTCIVGSSGCHLAHNLTKPIWEXDRFMTFQKEDARSILIRE
    EIVYHCFPTNRNAYVNLVKYKTLILAFKILRGRPRVYIISGAAYAVPPFYIGKLFQCKRK
    VYIVLEFPRIDKPTLTGKLVYFVTDKFTVQWEMKQVPRKINLGGIF"
    2249..2256
    2264..2746
    2264..2746
    /gene="cpsbG"
    2264..2746
    /gene="cpsbG"
    /function="beta-1,4-galactosyltransferase"
    /codon_start=1

```

```

/evidence=not experimental
/transl_table=11
/product="CpsIBK"
/protein_id="BAB21599.1"
/db_xref="GI:12697596"
/translation="M1PVTGTHGQQRRLKEVRLKGTGADIDGVFIQGYSPDF
ONQMSKPLSYDMSYMKAEVITHGDPATFNAAVSKKKTIVPRQDQGEHVN
HOVDPLKELFLKYEIDYILNISLENNIKENKISTSKVLSQNDPCSSFKDELKPLS
"
RBS
2806..2812
2842..3981
/gene="cpsIBH"
2842..3981
/gene="cpsIBH"
/function="putative CP polymerase"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="CpsIBH"
/protein_id="BAB21600.1"
/db_xref="GI:12697597"
/translation="MQLLLALLVLLICSSYNEKQKFLNNAIEFFIVYVYVSVI
LNSLSPFHFVIAAFNSIAGVSLFHYKNTNIELTLKSLFNALILCLG
FLYYAIVDFENVSLFGRNLGSDWINGMTHORAMAFEEYGNLIIPLIINIIYI
YIKORSSGMMLIGALSTIILPISGSGAGIIVLAVIILLNIIIVIKQITREFL
YVPIILILVILRPMLVSIYRNINRSGSSSESESLKDTVSIVLIDSLFLKGV
KELWMSDLPLSGSHYTGIFKTKGLPGLINVLIGFLILISITIKAKSDPYEIVG
SVLLSFPLALDIDGANWLIIFVTGLIDLENKQPYSQLKRES"
3964..3970
/gene="cpsIBH"
3961..4955
/gene="cpsIBH"
3961..4955
/gene="cpsIBH"
/function="beta-1,3-N-acetylglucosaminyltransferase"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="CpsIBH"
/protein_id="BAB21601.1"
/db_xref="GI:12697598"
/translation="MEKQILVNSIVIPINSEAYLKACVSVLQTHSLIEVLLINDS
TNSGEICDNLQKDRILVFKRNGGVSSARLGLDKSTGEITTFVDSDDVAPIT
EIMLKNLIEDADIAVDPDISNERDYKRRNFYKFNKNSLKKEFSKRVENITV
CTLYKSLIGNRPDENLKIGEDLFNCILQCEHCIVDTSSLYTVRIYKTSAN
OEENENSIDRTITFENISISIVPAKLANVYEAFLREKRCIKRMPELGSNIDKILQ
RSLPFQDKIYPTFYAVKYLISLKGSLSIYMKCSPLIYIKLDRFQKQ"
4995..5001
5008..5949
/gene="cpsIBH"
5008..5949
/gene="cpsIBH"
/function="beta-1,3-galactosyltransferase"
/codon_start=1
/evidence=experimental
/transl_table=11
/product="CpsIBH"
/protein_id="BAB21602.1"
/db_xref="GI:12697599"
/translation="MNYSTIMSYNEPLNVRDVSIEINQTLTDEFLIIVDPSRG
DKQPLREXSVNRIKILNEENIGLASSINAKVISEGVEYFERMADISPSFED
KQIRFHEENLDSATILILIDOKNIVYKORSNKIYTLNDRKLLNLSLIIHTW
KQKKVVDKLMGRDLPVVDYDPAIRGLADRFGLGNLVLQTLNENSGISQTKP
CQYKYSAILQDPFKESKSIDITKIVNFGYVVKRYTQDELSKVELKSTSTIRK
LYICLYFKSPVLRLLINDINILVRLKLFGEKQSD"
5928..5935
/gene="cpsIBH"
5942..6880
/gene="cpsIBH"
5942..6880
/gene="cpsIBH"
/function="putative sialyltransferase"
/codon_start=1

```

```

/evidence=not experimental
/transl_table=11
/product="CpsIBK"
/protein_id="BAB21603.1"
/db_xref="GI:12697600"
/translation="MTRKLIYVCHTLVHLICLYKEEISYNLIELIISSSIPDVNLEK
KLKSTNIIHILERSGSEBELSVLKDAGLSYKPSNCFIENDATPIGRLLIKGI
YVNIIEGLNCFYSISQCLMKYVKKYILHKKIOPGSPRCIGIENSLVNLKXP
RYKPIEIPREKLEFDNTEYOKEMALNLPKAVRSTISPSVLCITQPLSIDKEMSYN
NKIETSEQRNFYSIVNEYINKGVNVLNHPDVVDSKLPVELLPSVPMETIEL
MLNRFEGCIHSTALDFLCVCKKILTVLKLKIK"
6865..6870
/gene="cpsIBK"
6877..8277
/gene="cpsIBL"
6877..8277
/gene="cpsIBL"
/function="putative repeat unit transporter"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="CpsIBL"
/protein_id="BAB21604.1"
/db_xref="GI:12697601"
/translation="MTRKLLQSGRIYTLGNLVQGLAFITLPYTRVISTEYQGS
LYVAMNIMLPIGLQSGSLASARVKGEEFKSAGSPVGNIMPLIALLFLR
SLAPLVGSPSIFLWVCGSYANVVTFFQOYRLOQORSLANLISLANVSVALS
LFLPHSDDPLSLGVFPETITIGIYAPAYIYHXSFPNYPFRIIVTVSLPIF
HLHQLNLDRLMLNARISTKRVANSPGVSCMIIOIYINIMMIPFDPARK
ELKQLSTYISRYVILGVFLGVTVFPELAQIMGDDQSSVQFISLITVSVYK
LYTPVNIQFPFANVTNMPDITGLAAGVNLNLVLLPHVAYGAMAMTISYALLI
FHHVSKVYKYSVSVQYIILSGIYFVSAMNMFLNIVISWSJLIIIVYSIV
FOKVIDLSKRRRR"
8262..8269
/gene="cpsIBL"
8277..9302
/gene="neub"
8277..9302
/gene="neub"
/function="sialic acid synthetase"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="neub"
/protein_id="BAB21605.1"
/db_xref="GI:12697602"
/translation="MVIYIIAIGNHNNDINLAKKMDVAVSCGYDAVKFQFAEK
LSKPAKAYOKATGTADSOLEMTKRLSEFEYLSMRDYAISKGVTFSTPDEBS
LRLSTDMPIYKIPSGRITNLPLYEKIKGQOKVILSTGMAVMEIHQAVNIIQNG
TTDISIHCCTBYPTPYPSINLNLHTLKDKRPDLITIGYSDHSISGEVPIAAAGAE
VIEKFTLDTNMEVDFKASATPDILALVKGFPALINQALSRFKETIPDVEKKNIVA
RSVVALPKIKGDIYSIENTVVRPANGISPMWYDILGQADQDFEDVITDSRF
ENQLEL"
9367..9374
9379..9987
/gene="neuc"
9379..9987

```

Query Match 29.5%; Score 312; DB 1; Length 9987;
Best Local Similarity 57.5%; Pred. No. 2.2e-40;
Matches 608; Conservative 0; Mismatches 435; Indels 15; Gaps 2;

```

1  ATGAGTAATATATATATGCTGTAATTTGGTGCACCAATATGATGATAT 60
8271  AAGATATGCTTATATATATTCAGAGATGCTTGCATCAATATGAGATATATCT 8330
61  GCAAGGAAATGATATTAACCCAAAGAGCCCGGTGTAATGCGTAATAATCCAAACA 120
8331  GCAAAAAAATGATAGATGCTCCCGTCTTGCTGGTGGATGCTGTAATAATTCAGACT 8390
121  TTTAAGCTGATTAATTAATTCAGCTATGCAACCAAGGAGAGATCAATCAATAAACAAC 180
8391  TTTAAGCTGAGAACTTATTTCTAATTTGCTCCCAAGCTGATATATCAAAAGCAACT 8450

```

QY 181 ACAGGAAATTAGATCTCAGTTAGAAATGACAAAAGCTTGAAATGAGTATGACAT 240
 DB 8451 ACAGGAAACGACACACCTACTGATGATGATGATGATGATGATGATGATGATGAT 8510
 QY 241 TATTCATCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
 DB 8511 TATTCATCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8570
 QY 301 GAGGAGACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
 DB 8571 GAGGAGACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8630
 QY 361 GGTGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
 DB 8631 GGTGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8681
 QY 421 AAAATATCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 480
 DB 8682 AAAATATCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8741
 QY 481 TTTTAAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 540
 DB 8742 TTTTAAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8795
 QY 541 CCAAGCCCTTTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
 DB 8796 CCAAGCCCTTTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8855
 QY 601 AAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 660
 DB 8856 AAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8915
 QY 661 CCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720
 DB 8916 CCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8975
 QY 721 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 780
 DB 8976 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 9035
 QY 781 GTTGAAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840
 DB 9036 GTTGAAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 9095
 QY 841 ATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 900
 DB 9096 ATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 9155
 QY 901 GAAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 960
 DB 9156 GAAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 9215
 QY 961 TTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1020
 DB 9216 TTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 9275
 QY 1021 TTTCAAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1058
 DB 9276 TTTCAAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 9313

RESULT 14
 CUI1168X5 26335 bp DNA linear ECT 09-JUL-2002
 LOCUS CUI1168X5
 DEFINITION Campylobacter jejuni subsp. jejuni NCTC 11168 complete genome;
 ACCESSION AL139078 AL111168
 VERSION AL139078.2 GI:6968723
 KEYWORDS
 SOURCE Campylobacter jejuni subsp. jejuni NCTC 11168
 ORGANISM Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;

REFERENCE
 AUTHORS
 1. Campylobacteriaceae; Campylobacter.
 Parhill, J., Wren, B.W., Mungall, K., Kelley, J.M., Churcher, C.,
 Basham, D., Chillingworth, T., Davies, R.M., Feltwell, T., Holtroyd, S.,
 Jagsall, K., Karlyshev, A., Moule, S., Pallen, M.J., Penn, C.W.,
 Ogall, M., Rajandream, M.A., Rutherford, K.M., VanVleet, A.,
 Whitehead, S. and Barrall, B.G.
 The genome sequence of the food-borne pathogen Campylobacter jejuni
 reveals hypervariable sequences
 Nature 403 (6770), 665-668 (2000)
 JOURNAL
 MEDLINE
 PUBMED
 20150912
 10688204
 2. Bases 1 to 26335)
 Parhill, J.
 Direct Submission
 Submitted (09-FEB-2000) Submitted on behalf of the Campylobacter
 sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
 Hinxton, Cambridge CB10 1SA E-mail: parhill@sanger.ac.uk
 Notes:
 Details of C. jejuni sequencing at the Sanger Centre are available
 on the World Wide Web.
 (URL, http://www.sanger.ac.uk/projects/C_jejuni/).
 Location/Qualifiers
 1. 26335
 /organism="Campylobacter jejuni subsp. jejuni NCTC 11168"
 /mol_type="genomic DNA"
 /strain="NCTC 11168"
 /db_xref="taxon:192222"
 /complement(38, .1369)
 /gene="accC"
 /complement(38, .1369)
 /gene="accC"
 /EC_number="6.4.1.2"
 /EC_number="6.3.4.14"
 /note="Cj1290c, accC, probable biotin carboxylase (subunit
 of acetyl-CoA carboxylase (EC 6.4.1.2)), len: 443 aa;
 highly similar to many e.g. ACCC ECOLI biotin carboxylase
 (BC 6.3.4.14) (449 aa), fasta scores: opt: 1560 z-score:
 1741.0 E(): 0, 53.2% identity in 442 aa overlap, 67.0%
 identity in 445 aa overlap). Also similar to Cj1037c pycA (46.5%
 identity in 445 aa overlap). Contains P500866 and P500867
 Cardamoyl-phosphate synthase subdomain signatures 1 and 2,
 and Pfam match to entry PF00289 CPase_L_chain,
 Cardamoyl-phosphate synthase (CPase)"
 /codon_start=1
 /transl_table=11
 /product="biotin carboxylase"
 /protein_id="CAB73717.1"
 /db_xref="GI:6968724"
 /db_xref="GOA:O9PN09"
 /db_xref="SPTREMBL:O9PN09"
 /translation="MEIKSILILANNGEILALRLIKENGKALCVYSBNDALYLK
 YVADSLICIGKASSSESYLNIIPALIAAEADAPIPGGLSENFPEIICAKNHLK
 FPGSPYEAAMKMSDESKAKOMHAGVPYIPSDGALAGAAKALKEIGYVILKA
 AAGGGRGRVYENKDLKAYASAESEAMTIPGCTGMVYIIONPHIEVDIGDS
 FGVNIVHGRDCSMORHOKLIESPAIILDEKTRRLHETAIKAKAIGYBAGPFE
 FVNDKILDEFIEMNTRLOVHCSEVSGIDILISQIVASGIALPSGSLVNGHS
 IBCRLTVAESKTFLESPGKITXYIIPAGRNVMESHCVYDVSVPAYDNLGLVWA
 EDRNKALIAOKVALDELISGKITXDFILSMENMDPFINNNYDNYLNRH"
 /complement(47, .1357)
 /gene="accC"
 /note="Pfam match to entry PF00289 CPase_L_chain,
 Cardamoyl-phosphate synthase (CPase), score 640.70,
 E-value 8e-189"
 /complement(488, .511)
 /gene="accC"
 /note="P500867 Cardamoyl-phosphate synthase subdomain
 signature 2"
 /complement(866, .910)
 /gene="accC"
 /note="P500866 Cardamoyl-phosphate synthase subdomain
 signature 1"
 /complement(1371, .1826)
 misc_feature
 misc_feature
 misc_feature
 gene

```

CDS
/gene="accB"
/complement(1371..1826)
/gene="accB"
/BC_number="6.4.1.2"
/note="Cj1291c, accB, probable biotin carboxyl carrier
protein of acetyl-CoA carboxylase, len: 151 aa; similar to
e.g. BCCP-ECOLI biotin carboxyl carrier protein of
acetyl-CoA carboxylase (EC 6.4.1.2) (156 aa), fasta
scores: opt: 307 z-score: 329.6 E(): 4.6e-11, 38.9%
identity in 162 aa overlap, 41.4% identity to HP0371.
Contains Pfam match to entry PF00364 biotin_req_enuy,
Biotin-requiring enzymes"
/codon_start=1
/transl_table=11
/product="putative biotin carboxyl carrier protein of
acetyl-CoA carboxylase"
/protein_id="CA873718.1"
/db_xref="GI:6968725"
/db_xref="GOA:Q9PN08"
/db_xref="SPTREMBL:Q9PN08"
/translation="MTKEIKELVNLFAEANTSKIKIKQSGFEIELEPDMCCVDPAP
VCPVAPQPIVNVNEQPSQVSNRPSINSVGFYQAPSPGAPAFKVGSTV
KKGDPIATIEAKKIMNEIEABDCRIATILVADQPVFEGMPLFAVERL"
/complement(1383..1592)
/gene="accB"
/note="Pfam match to entry PF00364 biotin_req_enuy,
Biotin-requiring enzymes, score 104.20, E-value 2.5e-27"
1974..2534
/gene="dcd"
/BC_number="3.5.4.13"
/note="Cj1292, dcd, possible deoxycytidine triphosphate
deaminase, len: 186 aa; similar to many predicted dcd
proteins, and to DCD-ECOLI deoxycytidine triphosphate
deaminase (EC 3.5.4.13) (193 aa), fasta scores: opt: 160
z-score: 203.5 E(): 0.00049, 28.8% identity in 160 aa
overlap, 71.4% identity to HP0372. Contains Pfam match to
entry PF00692 dutPase"
/codon_start=1
/transl_table=11
/product="possible deoxycytidine triphosphate deaminase"
/protein_id="CA873719.1"
/db_xref="GI:6968726"
/db_xref="GOA:Q9PN07"
/db_xref="SPTREMBL:Q9PN07"
/translation="MGKADNMRIRMALEKHPCEANIGKGVSYGSLGYDIR
VVERKIFTNVNSIVDPKPFVEENVDRGCVCIYPAASPLARTIEYRKPQNVLA
ICLGSTVARGCIVAVTPFEGFSGHITELISNTTLPFAKIYANGIAQVLEQSDS
KCDITVKKRKGXQAQTGITLPRILK"
2046..2522
/gene="dcd"
/note="Pfam match to entry PF00692 dutPase, dutPase, score
3.80, E-value 0.0016"
2584..3588
/gene="Cj1293"
/note="Cj1293, possible sugar nucleotide
epimerase/dehydratase, len: 334 aa; similar to many e.g.
TR-Q45984 (EMBL:U27301) Caulobacter crescentus FLA1.
Protein (331 aa), fasta scores: opt: 1240 z-score: 1410.9
E(): 0, 57.8% identity in 325 aa overlap, CAPD_STRAID CAPD
protein (559 aa), fasta scores: opt: 529 z-score: 602.6
E(): 2.9e-26, 33.9% identity in 319 aa overlap, and
TR:069130 (EMBL:AF064070) Burkholderia pseudomallei
putative epimerase/dehydratase MBII (637 aa), fasta
scores: opt: 433 z-score: 493.6 E(): 3.4e-20, 30.5%
identity in 282 aa overlap, 64.6% identity to HP0840.
Contains PS00017 ATP/GTP-binding site motif A (P-loop)"
/codon_start=1
/transl_table=11
/product="possible sugar nucleotide epimerase/dehydratase"

```

```

misc_feature
/protein_id="CA873720.1"
/db_xref="GI:6968727"
/db_xref="GOA:Q9PN06"
/db_xref="SPTREMBL:Q9PN06"
/translation="MEFKKILLTGGSFGKTYKVLLENYXNKIILYSRDELKOP
EMASVPAPGARYIGIDVRKERSLAMPQVDFVTHAAMKHVLAENMECIKTI
HGMNVYDACPENYGVCKITLSDKACNPVNLGATLADKLPVAAVNTAGNCTRF
GVTRGVNVSSRGSVDPFKKLISSECKEIPIDPTWTRWIMISDQVLTVEFEY
HSGEIRLPKIPSKMIDTDLALALANSHKLTIGIRAGKHEIMISSDSHLTYFEFY
YALSPRIKPKNDPISNAGEXQKXGVDSPNNPLMASEKLELINTHEGP"
2617..2640
/gene="Cj1293"
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"
3530..4720
/gene="Cj1294"
3590..4720
/gene="Cj1294"
/note="Cj1294, probable aminotransferase, len: 376 aa;
similar to many members of the degT family e.g. SpEC BACSV
spore coat polysaccharide biosynthesis protein (389 aa),
fasta scores: opt: 586 z-score: 676.7 E(): 2.1e-30, 35.3%
identity in 385 aa overlap, and TR:088001 (EMBL:AJ007747)
Bordetella bronchiseptica putative amino-sugar
biosynthesis protein WBC (366 aa), fasta scores: opt: 482
z-score: 558.1 E(): 8.7e-24, 28.8% identity in 330 aa
overlap, 44.8% identity to HP0366. Contains Pfam match to
entry PF01041 DegT_Dnrd_EryC1, DegT/Dnrd/EryC1/StxS
family"
/codon_start=1
/transl_table=11
/product="putative aminotransferase (degT family)"
/protein_id="CA873721.1"
/db_xref="GI:6968728"
/db_xref="SPTREMBL:Q9PN05"
/translation="MTYSRNTQSGSIDITKALKDLEILGKXVNEFRELCEWNG
VRAKCVNSATSAHLALVYALGVEKVLITPLFAATANAALMAKQVEIDIKNG
NIDERKLEARLKSEENIGALISVDVFGNSVEMDEINLTKVNIPLIDDSHLAGL
DYSEKVGKADLSIFSPHPVPIITTFEGGAVSNBELIDIKILRSHGIYKXLMDS
IMVELGVNYSRLDVALGINOLKDLHNEKREELINFIYKEPKKPYSTTIXIYQ
KKSSRHLYPILLFPFPGQKEELFESLIHAGIYGVHVKPFYEFKVLGEIKLQV
ADNRKFAISLIPCHQENLMDAKRVKDTLPSILKKVKGYGCG"
3608..4681
/gene="Cj1294"
/note="Pfam match to entry PF01041 DegT_Dnrd_EryC1,
DegT/Dnrd/EryC1,
28.4%; Score 300.6; DB 1; Length 26335;
Best local similarity 57.7%; Pred. No. 6,4e-39;
Matches 581; Conservative 0; Mismatches 414; Indels 12; Gaps 2;
Query Match
1 ATGAGTAATATATATATGCTGCTGAAATGCTGCAACCTAATGCTAGTGTGATTT 60
DB 31867 ATGAAAAAACTTAAATATGCTGCAAGACAGAGTAAATGCAATGAGATTAAATT 31926
61 GCAAGAGAAATGATATTAAGCAAGAGGCGGCTGTAATGCAAGTAAATTCGAAACA 120
DB 31927 GCTAAAGATTAAAGAAATGCGACACATTCAGAGCTGATTTGTTAAATTTCAAGT 31986
121 TTTAAAGCTGATTAATTAATTAATGCTATTCACCTTAAGGCAAGATCAATTAAGAAC 180
DB 31987 TTTAAAGCAAAACTGATGACGACAAAGGCTTAAAGACCTTACACCTTAAAGC 32046
181 ACAAGAAATTAAGATTCGATTAAGATGACAAAGAAAGCTTGAATGAAGATGACAT 240
DB 32047 ACGGTAATGATGAAGTCAAGCTTCAATGTGCAAAAATCGAATCGAATCTTAAAGT 32106
241 TATCTCATCTAATGATATGCAAGTCAAGTAAATTAATTAATGATGTTTCTACCCCTTT 300
DB 32107 CATAAAGGCTTATCTTACGCAAAAAGCAATATGCGCTTCTTCTTACCTCTTTT 32166
301 GACGAAAGCTCTATGATTTTGTGACATCTTGAACAAAGAAATATGGAATACCTCTTCA 360
DB 32167 GATCTGAAGGTGATGATCTTTTAAATGATGAGCTTAAAAATCTTTAAATTCGAAAGC 32226

```

QY 361 GGTGACTATTGATTTACCGTATCTTGAAATAATAGCCAGCTTCCGATCCCTGATAG 420
 DB 32227 GGAGAAATGACCAATCTACCTCTATCTTAAATAATGCAAGCT-----TATATA 32277
 QY 421 AAAATATCATATCAACAGGAATGGCTACATTTGATGATGATTAACAGCTGTCTTAT 480
 DB 32278 AAAATATCTTCTTCAACAGCAATGGCAAAATTTAGGCAAGTATAGCAAGCTTTAAATG 32337
 QY 481 TTATATAATATAAAGTTCGGGTGATATATATATATATATATATATATATATAT 540
 DB 32338 CTTTGCATAAATGGCGCTAAGCCCAAAATATATCTTATATATATATATATATAT 32397
 QY 541 CCAAGCCCTTTGAGATGTAACCTTATATGATTAATGATTTGAAAAACCTTCCCT 600
 DB 32398 CCGGCCCTTTTATGTAAGTCAACCTTAAAGCAATGCAAGCTTAAAGACGCTT---T 32454
 QY 601 AAGATATATATAGCTTCTGTGATATATATATATATATATATATATATATAT 660
 DB 32455 AAGCTGATGATGATATATATATATATATATATATATATATATATATATAT 32514
 QY 661 CCTATATGAAAT 720
 DB 32515 GCCTTGGAGACTGTGTGATATGAAATATATATATATATATATATATATATAT 32574
 QY 721 GATCATTTGGCCCTCATATGAACTGATGAACTGATGAACTGATGAACTGATG 780
 DB 32575 GATCATTTAAACCAAGCCCTGATGAACTGATGAACTGATGAACTGATGAACT 32634
 QY 781 GTTGAAATAATCTTATAGTTCATAATATATATATATATATATATATATATAT 840
 DB 32635 ATACAAATGCAATGATGATGATGATGATGATGATGATGATGATGATGATG 32694
 QY 841 ATCTGACCAAGAAATGCTATATATATATATATATATATATATATATATATAT 900
 DB 32695 AACTATGTAAGAAATGCTATATATATATATATATATATATATATATATATAT 32754
 QY 901 GAAAAAAT 960
 DB 32755 GAAAGAAATCTAT 32814
 QY 961 TTATGCGTAAATATGCAAGCAAGCTTATATATATATATATATATATATAT 1007
 DB 32815 TTTTGGTAAATGCTATATATATATATATATATATATATATATATATATAT 32861

RESULT 15
 AY102621
 LOCUS 13390 bp DNA linear ECT 08-SEP-2003
 DEFINITION Campylobacter coli Cj1314c-like protein (Cj1314c) gene, partial cds; Cj1315c pseudogene, complete sequence; Cj1316c pseudogene, complete sequence; and NeuB3 (neuB3), Cj1319-like protein, Cj1320-like protein, Cj1324-like protein, Cj1325-like protein, Pmc (pmc), PtmB (ptmB), PtmC (ptmC), PtmF (ptmF), and PtmA (ptmA) genes, complete cds.
 VERSION AY102621 AF195053 U25992
 KEYWORDS
 SOURCE Campylobacter coli
 ORGANISM Campylobacter coli
 Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales; Campylobacteraceae; Campylobacter.
 REFERENCE
 AUTHORS Query, P., Doig, P., Alm, R.A., Burr, D.H., Kinsella, N. and Trust, T.J.
 TITLE Identification and characterization of genes required for post-translational modification of Campylobacter coli VC167 flagellin
 JOURNAL Mol. Microbiol. 19 (2), 369-378 (1996)
 MEDLINE 96423180
 PUBMED 8825781
 REFERENCE 2 (bases 1 to 13390)
 AUTHORS Logan, S.M., Kelly, J.F., Thibault, P., Ewing, C.P. and Query, P.
 TITLE Structural heterogeneity of carbohydrate modifications affects

JOURNAL serospecificity of Campylobacter flagellins
 MEDLINE Mol. Microbiol. 46 (2), 587-597 (2002)
 PUBMED 22294811
 12406231
 REFERENCE 3 (bases 1 to 13390)
 AUTHORS Query, P.M., Doig, P., Alm, R.A., Burr, D.H., Kinsella, N. and Trust, T.J.
 TITLE Direct Submission
 JOURNAL Submitted (02-MAY-1995) Enteric Diseases, Naval Medical Research Center, 503 Robert Grant Ave, Silver Spring, MD 20910, USA
 REFERENCE 4 (bases 1 to 13390)
 AUTHORS Query, P., Ewing, C.P., Moran, A.P. and Trust, T.J.
 TITLE Direct Submission
 JOURNAL Submitted (14-OCT-1995) Enteric Diseases, Naval Medical Research Center, 503 Robert Grant Ave, Silver Spring, MD 20910, USA
 REFERENCE 5 (bases 1 to 13390)
 AUTHORS Logan, S.M., Kelly, J.F., Thibault, P., Ewing, C.P. and Query, P.
 TITLE Direct Submission
 JOURNAL Submitted (13-MAY-2002) Enteric Diseases, Naval Medical Research Center, 503 Robert Grant Ave, Silver Spring, MD 20910, USA
 REFERENCE Sequence updated by submitter
 COMMENT On or Before Jun 2, 2003 this sequence version replaced gi:1049049, gi:11095580.

FEATURES
 source
 1. 13390
 /organism="Campylobacter coli"
 /mol_type="genomic DNA"
 /strain="VC167"
 /db_xref="taxon:195"
 complement(1..558)
 /gene="Cj1314c"
 complement(1..558)
 /gene="Cj1314c"
 /note="probable cyclase"
 /codon_start=1
 /transl_table=11
 /product="Cj1314c-like protein"
 /protein_id="AAW76271.1"
 /db_xref="GI:31324510"
 /translation="MLKRIIPCVLKNLSQVLSIEFQDFTIGLSTPIRYANRV
 DELIITDIDSKSGKIDPSIEDIAKQCMPLTIGGKITLEDIKIINIGADKISIN
 SKALHMDPIISKANRSGQICVCSIDYRKRGQCYVDKSLKRSPLALAEYRK
 GAGELITVSDDEGRKGYDLEMLK"
 complement(559..1163)
 /gene="Cj1315c"
 /pseudo
 complement(1160..2297)
 /gene="Cj1316c"
 /note="synonym: psea; involved in synthesis of psea"
 /pseudo
 2447..3478
 /gene="neuB3"
 2447..3478
 /gene="neuB3"
 /note="required for flagella synthesis"
 /codon_start=1
 /transl_table=11
 /product="NeuB3"
 /protein_id="AAW76272.1"
 /db_xref="GI:31324511"
 /translation="MOIGNFNTDKVFIILASAHHTSLLENALOSIKAAKAGADAI
 KICATYTPSITNSKEDIKIGLNDKRYELYSAKTPEYMSQIFETTONBGL
 CRRSPFAKEDVEFLKRPDLIAKRIASFPANDENPFLIAKKKPTVSGIATBEFL
 KICEIPEKERNPDLFLKSTIYPIRAIDMKIGVLSKEKNVRYGSDHSRGFLAP
 VNAVALGAVIEIKHMDKSISSDSKSLDDEFRKAVADAYRQESALGDKDLDB
 KALKNRVPAKSLYAKMDIKKGMFSEENKVSVPSPGLPKPYQELGKASKDIIEFG
 DALKQDDEK"
 3475..4446
 /note="ORF1"
 /codon_start=1
 /transl_table=11
 /product="Cj1319-like protein"
 /protein_id="AAW76273.1"

/db_xref="GI:31324512"
 /translation="MRNLITVAGADPSITSHCESLIVKKGKGVKRALISQNSPFWKHLE
 KSPFLKMEVLSGLDRDSFCEKTKTNIDALFHDALAIAPYSTAPADYVDIVNGI
 LAMLEAAKKNESISHFHTSTSEYGTAVYIDEKHPLQSPSASASIAADWALSY
 YNSFNLNIAAPFNITGPPOSARAIPTIITOLISGAKELKGLDLSFKDLNPLVLT
 CEGFSLINKHKGFEVNIIGSVEMOEVLDILOKIDLSYKCIODEQRALPNSSEV
 FRLCDKAKIKKATWQSKISLEELRQSIYFKNLENTSELYNN"
 4439..5581
 /note="ORF2"
 /codon_start=1
 /transl_table=11
 /product="Cj1320-like protein"
 /protein_id="AAW76274.1"
 /db_xref="GI:31324513"
 /translation="MEKEISIPKSLFKENKNIALHEPCFIENKKYLLICTDNGFVS
 VGEFVREBALEKTKAFVATNTGTAAHIALLANGIDENCEVITQSSISPAATAN
 AIAVAGARAFIDIDENILSGPKALQNFLENHSTOKGFYTKITRAIRACVAMHT
 FGLSADIERLCICQNYNIFLEDAALAGSTYKNAKALGTGKGLISFNQKLIITGG
 SGAILCDENILANAKHLSTAKIPHEVEDHDIIGYVRLCNINAIILAGLEKLD
 FLEKRELAQIYKSEFTNKKACFIDERNKSKSNFALALCPKQNLNATFLQBLE
 NITFARPVWKSPLKPGQWQKDLNLTQKLEKLVNLPSSVRLN"
 5822..6793
 /note="ORF3"
 /codon_start=1
 /transl_table=11
 /product="Cj1324-like protein"
 /protein_id="AAW76275.1"
 /db_xref="GI:31324514"
 /translation="MPTRPQIKFKPKDKRKNISACINHKKEKIDYKARFKLENT
 CKYRMNGKEBYDCAIAVSGKDSHYOHIMKEKLGAPILFSVEDNPTMGARKN
 LNLSEITGCHLISLSPDKTKQKWLKTFEYGVPTWIDLIYSFPGALKTPT
 LLVGENSYVEYGGDDKEIPSAKAIPEWGAIDIDIEFIDDEIKENIDLPEPSA
 DMLNKDPIYLSYFIKMSYSYIIPAKSGFSDLOMDRSHCANPQIITIGYIILH
 AMWKPFGHMAASDYASFRVYGMWSREBAKVIKQMDHKLNNKN"
 6932..7612
 /note="ORF4"
 /codon_start=1
 /transl_table=11
 /product="Cj1326-like protein"
 /protein_id="AAW76276.1"
 /db_xref="GI:31324515"
 /translation="MNSICTYTKYDENNGVGLYDPDSVIRFYERILKYKLGITSN
 LLDPGGNSVHKFELDKRFPQIDIVPSLOKMDAKIKISGGGDCIKIENSSIKG
 LFDKMDIVFANQSLYIPKTELKONILFEYCLNQGILPTMSEKMYSHSEKX
 NENGLRKSIOGRINLETIYIHPAKSIDELQELKPEALFDGIDPINTFNPESAH
 FYIIGIKR"
 7654..8658
 /gene="ptmc"
 /note="involved in synthesis of glycosyl modifications of
 flagellin"
 /codon_start=1
 /transl_table=11
 /product="ptmc"
 /protein_id="AAW76277.1"
 /db_xref="GI:31324516"
 /translation="MKKALIIABAGVNHNDLNAKLIIVAKSGADPYKRSKAS
 LCVSKNAKAAAYQLKTAKDESQLEVIKLELDFNHOILISHAQCGIATSTAPDL
 ESIELDDJLEVFKIPSGIITNLPLYKQIADINKITLISTESLQGIIEALVELYCK
 EGTORANITILICTEYPAFVFNVLKQIDKMAENLWGVSGDNGHILHIAL
 GASVIERFTLDKMGSPDKASLEBDELOELCTKRELSALGQGIKXASSEKNI
 EIAKSLVAKKIKKGEITSEENLTKRPASGISARVDEYIKKXSKDYEDDLIHE
 "

/protein_id="AAW76278.1"
 /db_xref="GI:31324517"
 /translation="MSKRKICIVASRAEYLLRNLCLEIIONDELOIITATGALS
 PREGLYTKIEKRFKAKIKIPIILASDDKSLCKMAALNIGSEBAPESLKPDIYVL
 GDEYEMLSVAVSLAMAPLPAHICGSELITGALIDBSIRHSISMAHLFPVQTYNR
 LDLOGEKSNVNSVLASTTMKNINLENGELAKYLEPDLQNYILITYHPTNLKX
 NSKREIDILIKTIDELNNAHLPTKNADENGLLINSELDYCKSKPRKALPDNQS
 OKYLSIKIKAMIGNSSGISSEPFKPCINIGSRQKRLRTQNIIDCFENLDA
 FEKLSKDFKQALHFFKNPYENDKNPILIKOLKNAAMDITLIQKNFIDLRNLGYK"
 10182..10844
 /gene="ptmc"
 /note="involved in synthesis of glycosyl modifications of
 flagellin; nucleotidyl transferase"
 /codon_start=1
 /transl_table=11
 /product="ptmc"
 /protein_id="AAW76279.1"
 /db_xref="GI:31324518"
 /translation="MAGLGTRUKELTQDTPPKMLKYGKRPILASITVQIKREPNF
 IFCVNYKQIIEYFKKGNVYKFIYIKERKLGAGLSLIKQDKESFVWAKDI
 LTELNFELIKRHSKALMSVCRBEQOIPEGVTSQNGFTITITEKVFQKFLVSA
 GIVCEPEMELINKNEYILMPFLIKLAKQKGVNTYLIHDYIMIDIGRLDEPVMKED
 FK"
 10841..11746
 /gene="ptmc"
 /protein_id="AAW76280.1"
 /db_xref="GI:31324519"
 /translation="MAGLGTRUKELTQDTPPKMLKYGKRPILASITVQIKREPNF
 IFCVNYKQIIEYFKKGNVYKFIYIKERKLGAGLSLIKQDKESFVWAKDI
 LTELNFELIKRHSKALMSVCRBEQOIPEGVTSQNGFTITITEKVFQKFLVSA
 GIVCEPEMELINKNEYILMPFLIKLAKQKGVNTYLIHDYIMIDIGRLDEPVMKED
 FK"
 10841..11746
 /gene="ptmc"

Query Match 27.88; Score 294.2; DB 1; Length 13390;
 Best Local Similarity 57.38; Pred. 1.4e-37;
 Matches 577; Conservative 0; Mismatches 418; Indels 12; Gaps 2;
 1 ATGAGTATATATATATCGTTGCTGAATTCGTTGCAACCAATAGTAGTGGATTT 60
 DB 7654 ATGAAAAAGCTTTATCATATGACAGACAGGGCTTAATCAATGAGATTAAATTTG 7713
 61 GCAGAGAAATGATATTAAGCCAAAGAGCGGCTGTATGCAATTAATCCAAACA 120
 DB 7714 GCCAAAAAGCTGATGCAAGTTGCTGCCAAAGTAGAGTATTTGTTAAATTTCAAGT 7773
 121 TTTAAAGCTGATTAATTAATTTACGATTTGACCTTAAGGAGATCAATTAATAAAC 180
 DB 7774 TTCAAGCGCAACTTGGGTGAGTAAATACTCTAAAAAGCAGCTTACGCTAAAAACC 7833
 181 ACAGAGAAATTAATCTCAGTTGAAATGCAAAAAAGCTTGAATGAGATGACGAT 240
 DB 7834 ACAGCTAAAGATGAAGCAATGAAATGTTAAAAACTAGACCTTGACCTTAATGCA 7893
 241 TATCTCATCTAATGAAATGACAGTCACTTAATTAATTAATGATTTTGTACCCCTTT 300
 DB 7894 CACGAGCTTTAATCTCATGCTAATCAACAATGAGTATGCTTTCTTCAACGCTTTT 7953
 301 GACGAGACTCTATGATTTTATGACATCTTTGAAACAAAATGTAATGAAAAATCCCTCA 360
 DB 7954 GATTAGAAAGATTAAGACCTTTTATGATGATTAAGACTTGAAGTTTAAAAATTCAGAC 8013
 361 GGTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 420
 DB 8014 GAGGAATCACAAAATTTGCTTATCTAATAAAAAATTTGCAAAATTTA-----ATAAA 8064
 421 AAAATTAATCAATCAACGAGATGCTACTATGATGATGATTAATTAATTAATTAATTA 480
 DB 8065 AAAATCAATCACTCCACAGGAATGCAAAATTTGAGCGAGATGAGCGGCTTTAGAAAT 8124
 481 TTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 540
 DB 8125 CTTTGCAAGAGAGCACCACCAAGACCAATTAATCACTTCACTTCACTGACATGCAATAC 8184
 541 CCAGGCGCTTTGAGATGTAACCTTAATGCTATTAATTAATTAATTAATTAATTAATTA 600
 DB 8185 CCGGACCTTTAATGAGTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 8241

```
QY 601 AAGATTAACATAGGCTTCTCTGATCTCTAGGCGTTTATGACGCTATGCGCGGTG 660
Db 8242 AATCTGATGTAGGCTATTCAGATCAACCAAGGATATACATATTTCTTTGACCGCTATC 8301
QY 661 CCTTATGAAATTAATCTTTATGAAAAACATTTCACTTTAGATTAATCTATGTCGCGCA 720
Db 8302 GCCTTAGGGGCAAGCGTATAGAAAAGCATTTTACCTTAGATTAATAATATGTCAGGACCT 8361
QY 721 GATCATTGGCCCTCAATAGAACCTGATGAACCTGAACATCTTTGATTTGGGGTCAAGTGT 780
Db 8362 GATCATTAAGCGCTCTTTAGAACCTGATGAGCTTCAAGAACTTTGCACAAAATTTAGAGAA 8421
QY 781 GTTGAATAATCTTTAGGTCATAAATAGTAAAGTGTACAGCTTCAGAAAGGAAGATAAA 840
Db 8422 ATCGAAGTGTCTTTAGGTGATGGCATAAACAGCAAGCAAGCAAGCAAGCAAGCAAGCA 8481
QY 841 ATGTAGCAAGAAAGTCTATTATAGCTAAACAGAGATAAAAAGGTGAGGTTTTTCA 900
Db 8482 GAATTTGCAAGAAAAAGCCTAGTAGCTAAAAAGAAAAATCAAAAAAGAGAAATTTTAGC 8541
QY 901 GAAAAAATATTAACAACAAAAAGCCTGTATATGTATCCGATGCGATGAGTGTAAAT 960
Db 8542 GAGGAAATCTTACACAAAAAGCCTGTATATGTATATGCGATGCGCTTATGATGAG 8601
QY 961 TTAATGGGTAAATTTGCAAGCAAGCTTTATTTCCAGATGAATTAAT 1007
Db 8602 TATTTGGGCAAAAAAGCCAGTAAGATTGAAGAAAGATGAGCTAAAT 8648
```

Search completed: May 11, 2004, 17:15:55
Job time : 4316.94 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 10, 2004, 04:14:30 ; Search time 456.589 Seconds
(without alignments)

9853.151 Million cell updates/sec

Title: US-09-930-440b-7

Perfect score: 1059
Sequence: 1 atgagatataatataatcgt.....aatgagacaaataatcattg 1059

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_28Jan04:*
1: geneseqn19808:*
2: geneseqn19908:*
3: geneseqn20005:*
4: geneseqn20011:*
5: geneseqn20015:*
6: geneseqn20022:*
7: geneseqn20033:*
8: geneseqn20038:*
9: geneseqn20039:*
10: geneseqn20045:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB ID	Description
1	1059	100.0	1059	AAA50570	Aaa50570 Escherich
2	1059	100.0	1059	AAH28459	Aah28459 Nucleicid
3	1059	100.0	1059	AAH53994	Aah53994 DNA encod
4	326.4	30.8	17276	ACA64723	ACA64723 Streptoco
5	324.8	30.7	110000	ABW71527_11	Continuation (12 o
6	321.4	30.3	1023	ABW71527_11	Continuation (12 o
7	317.2	30.0	26281	AAZ60929	AAZ60929 Nucleicid
8	163	15.4	110000	AAV21309_04	Continuation (5 of
9	107.6	10.2	1050	ABA99765	ABA99765 N. mening
10	107.6	10.2	65632	AAA81502	AAA81502 N. mening
11	107.6	10.2	110000	AAA81490_00	AAA81490 N. mening
12	107.6	10.2	349980	AAV21544	AAV21544 Neisseria
13	105.2	9.9	11474	AAV21544	AAV21544 LPS core
14	105.2	9.9	11474	ABH13665	ABH13665 Campyloba
15	100	9.4	1000	ABK37785	ABK37785 DNA seque
16	91.4	8.6	999	AAH91405	AAH91405 N. mening
17	90	8.5	984	ABV72354	ABV72354 Nucleicid
18	70.4	6.6	1123	AAH14531	AAH14531 H. pylori
19	64.6	6.1	1170	AAH14557	AAH14557 Human cDN
20	63	5.9	1080	AAH37763	AAH37763 Human gly
21	63	5.9	1080	AAH28458	AAH28458 Nucleicid
22	63	5.9	1080	AAH53993	AAH53993 DNA encod
23	63	5.9	1146	AAA50569	AAA50569 Human sta

24	63	5.9	1230	AAI59142	AAI59142 Human pol
25	63	5.9	1230	ADB49125	ADB49125 Novel hum
26	63	5.9	1238	AAH31141	AAH31141 DNA encod
27	63	5.9	1243	AAI60928	AAI60928 Human pol
28	63	5.9	1268	ABL90630	ABL90630 Human pol
29	61	5.8	1671	AAQ24134	AAQ24134 50 KD sub
30	57.2	5.4	410	AAH35557	AAH35557 Human col
31	52.8	5.0	6394	AAH31374	AAH31374 Signal tr
32	52.8	5.0	6394	AAH31374	AAH31374 Human gen
33	52.8	5.0	34548	ABL70603	ABL70603 Chemical
34	52.6	5.0	5152	ABL92306	ABL92306 Human pol
35	52.6	5.0	5152	ABL92373	ABL92373 Human pol
36	51	4.8	514	AAH11876	AAH11876 Human cDN
37	50.4	4.8	611	ABX65932	ABX65932 Helicobac
38	50	4.7	547	ABX65406	ABX65406 Helicobac
39	49.6	4.7	6775	ABQ67159	ABQ67159 Human arg
40	49.4	4.7	6061	ABL312141	ABL312141 Human imm
41	49.4	4.6	3123	ABL28094	ABL28094 Droscophi
42	48.6	4.6	583	ADA72369	ADA72369 Rice gene
43	48.4	4.6	110000	ABA92787_0	ABA92787 Buchnera
44	48	4.5	11650	AAH46755	AAH46755 Tumour su
45	47.8	4.5	1431	AAZ37082	AAZ37082 DNA seque

ALIGNMENTS

RESULT 1	AAH50570	AAH50570 standard; cDNA; 1059 BP.
XX	AAH50570;	
AC	AAH50570;	
XX	19-DEC-2000 (first entry)	
DT		
XX	Escherichia coli stalic acid synthetase cDNA.	
DE		
XX	Stalic acid synthetase; neub gene; stalylation; glycoprotein;	
KM	plasmidogen; transsterin; chylotropin; Na ⁺ ,K ⁺ -ATPase; ss.	
XX	Escherichia coli.	
OS		
XX	Key	Location/Qualifiers
PH	CDS	1..1041
FT		/*tag= a
XX	W0200052135-A2.	
XX	08-SEP-2000.	
PD		
XX	01-MAR-2000; 2000MO-US005313.	
PF		
XX	02-MAR-1999; 99US-0122582P.	
PR	08-DEC-1999; 99US-0169624P.	
XX	(HUMA-) HUMAN GENOME SCI INC.	
PA	(UYJO-) UNIV JOHNS HOPKINS.	
PA	(UYWY-) UNIV WYOMING.	
XX		
PI	Betenbaugh MJ, Lawrence S, Lee YC, Jarvis D, Coleman TA.	
XX	WPI; 2000-572178/53.	
DR	P-PSDB; AAY96102.	
XX		
PF	Recombinant production of stalylated glycoproteins using cells in which	
PT	the expression of enzymes, e.g. stalic acid synthetase, involved in the	
ET	stalylation reaction has been altered.	
XX		
ES	Disclosure; Page 106-108; 144pp; English.	
XX		
CC	The present sequence is that of the Escherichia coli neub gene encoding	
CC	stalic acid synthetase (see AAY96102). The gene was used to identify the	
CC	corresponding human SAS gene (see AAA50569). The invention provides	

OY	1	ATGAGTAATATATATATGCTTGGCTGAATTTGGTTGCCAACCATTAAGTAAAGTTGATTT	60
Db	1	ATGAGTAATATATATATGCTTGGCTGAATTTGGTTGCCAACCATTAAGTAAAGTTGATTT	60
OY	61	GCAAGAGAAATGATATTTAAAGAGCCAAAGAGCCGGTGTAAATGACGTAATTTCCAAACA	120
Db	61	GCAAGAGAAATGATATTTAAAGAGCCAAAGAGCCGGTGTAAATGACGTAATTTCCAAACA	120
OY	121	TTTAAAGCTGATAAATTAATTTACGCTATTTGACCTAATAGGCGAGAGTATCAAAATPAAAAC	180
Db	121	TTTAAAGCTGATAAATTAATTTACGCTATTTGACCTAATAGGCGAGAGTATCAAAATPAAAAC	180
OY	181	ACAGAGAAATTGAAATCTCAGTTAGAAATGACAAATAAAGCTTGAATTAAGTATACGAT	240
Db	181	ACAGAGAAATTGAAATCTCAGTTAGAAATGACAAATAAAGCTTGAATTAAGTATACGAT	240
OY	241	TATCTCCATCTPAAAGAAATAGCAGTCAGTTAAATTTAGATGTTTTTTCTAACCCCTTTT	300
Db	241	TATCTCCATCTPAAAGAAATAGCAGTCAGTTAAATTTAGATGTTTTTTCTAACCCCTTTT	300
OY	301	GACGAAGACTCATTTGATTTTTCAGACTTTTGAACAACAAAATATGAGAAATCCCTTCA	360
Db	301	GACGAAGACTCATTTGATTTTTCAGACTTTTGAACAACAAAATATGAGAAATCCCTTCA	360
OY	361	GGTGAAGTATTAATTAATTTACCGTATCTTGAAAAATAGCCAAAGCTTCCGATCCCTGATAG	420
Db	361	GGTGAAGTATTAATTAATTTACCGTATCTTGAAAAATAGCCAAAGCTTCCGATCCCTGATAG	420
OY	421	AAATATATCATATCAACAGAAATGCGTACTATTTGATGAGATTAACAAGCTGTTCTCAT	480
Db	421	AAATATATCATATCAACAGAAATGCGTACTATTTGATGAGATTAACAAGCTGTTCTCAT	480
OY	481	TTTATPAAATATPAAAGTTCCGGTGGTAAATTTACAAATTTCACTTGCAATCTGAATAT	540
Db	481	TTTATPAAATATPAAAGTTCCGGTGGTAAATTTACAAATTTCACTTGCAATCTGAATAT	540
OY	541	CCAAAGCCCTTGGAGAGTAAACCTTAATGCTATTAAGATTGGAATAAACAATTCCT	600
Db	541	CCAAAGCCCTTGGAGAGTAAACCTTAATGCTATTAAGATTGGAATAAACAATTCCT	600
OY	601	AAAGATTAACATAGGCTTCTGTATCAATTTAGCGGGTTTTATGCAAGCTATTTGGGCGGTG	660
Db	601	AAAGATTAACATAGGCTTCTGTATCAATTTAGCGGGTTTTATGCAAGCTATTTGGGCGGTG	660
OY	661	CGTTATGGAATTAACCTTTATGAAAAACATTTCACTTTGATPAAATCTATGTCTGGCCCA	720
Db	661	CGTTATGGAATTAACCTTTATGAAAAACATTTCACTTTGATPAAATCTATGTCTGGCCCA	720
OY	721	GATCATTTGGCCTCAATAGAACCTGATGAACCTGAACATCTTTGATTTGGGGTCAGGTGT	780
Db	721	GATCATTTGGCCTCAATAGAACCTGATGAACCTGAACATCTTTGATTTGGGGTCAGGTGT	780

RESULT 4
 ID ACA64723 standard; DNA; 17276 BP.
 AC ACA64723;
 DT 18-JUN-2003 (first entry)
 DE Streptococcus capsular polysaccharide gene.
 KM Superantigen; *de* gene; *Sag*; staphylococcal enterotoxin; tumour; cancer;
 KM apoptosis; gene therapy; mammalian cell receptor; cyclostatic; APC;
 KM tumour associated lipid; energy; T cell; antigen presenting cell; APC;
 KM tumoricidal immunocyte; antitumour.
 OS Streptococcus sp.
 PM US2002177551-A1.
 PD 28-NOV-2002.
 PF 30-MAY-2001; 2001US-00870759.
 PR 31-MAY-2000; 2000US-0208128P.
 PA (TERM/) TERMAN D S.
 PI Terman DS;
 DR WPI; 2003-361759/34.
 DR P-Psdb; ABU79115.
 PT A mammalian cell receptor, useful in the treatment of cancer by binding
 PT to tumor associated lipids where the binding induces energy or apoptosis
 PT in T cells and antigen presenting cells.
 PS
 PS Example 2; Page; 167p; English.
 XX
 XX The invention relates to a mammalian cell receptor, useful in the
 XX treatment of cancer, which binds to tumour associated lipids and induces
 XX energy or apoptosis in the T cells and antigen presenting cells (APCs).
 XX Also included are a mammalian cell useful in the treatment of cancer
 XX where the receptor which binds tumour associated lipids and induces
 XX cellular inactivation or death is deleted or functionally deactivated,
 XX producing (M1) a tumoricidal immunocyte population in vivo in a mammal
 XX (by allowing tumour associated lipids to contact immunocytes in which
 XX receptors for immunosuppressive fatty acids, ceramides, gangliosides,
 XX sphingolipids, glycosphingolipids, phosphosphingolipids, gangliosides,
 XX sialylated glycans, lipopeptides and proteoglycolipids are inactivated or
 XX deleted), a construct useful in the treatment of cancer comprising a
 XX superantigen (Sag) nucleotide inserted into a virus, a mammalian T cell


```

QY 1 ATGAGTATATATATATGCTGCTGAATGCTTGCACCAATATGATGCTGATAT 60
DB 37197 AAGATTAAGGTTATATATATGACAGATGGTTCACATCAATATGAGATATATCTT 37138
QY 61 GCGAGGAAATGATATTAATAAGCCAAAGAGCGGCTGTTATGACGTAATAATCCAAACA 120
DB 37137 GCGAAAAAATGATGATGCTGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 37078
QY 121 TTTAAGCTGATATATATATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 180
DB 37077 TTTAAGCTGATATATATATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 37018
QY 181 ACAGAGATATGATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 240
DB 37017 ACAGAGATATGATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 36958
QY 241 TATCTCCATCTATGATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300
DB 36957 TACTTGAATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 36898
QY 301 GACGAGATCTATGATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360
DB 36897 GATGAGATCTATGATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 36838
QY 361 GGTGATATGATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
DB 36837 GGTGATATGATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 36787
QY 421 AAAATATATATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 480
DB 36786 AAAATATATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 36727
QY 481 TTTAATATATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
DB 36726 TTTAATATATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 36673
QY 541 CCAAGGCTTATGATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
DB 36672 CCAAGGCTTATGATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 36613
QY 601 AAGATATATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
DB 36612 AAGATATATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 36553
QY 661 CTTATGATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720
DB 36552 CTTATGATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 36493
QY 721 GATGATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
DB 36492 GATGATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 36433
QY 781 GTTGAATATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840
DB 36432 GTTGAATATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 36373
QY 841 ATGATGATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900
DB 36372 ATGATGATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 36313
QY 901 GAAAAATATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960
DB 36312 AATGATATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 36253
QY 961 TTTATGATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020
DB 36252 ATGATGATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 36193
QY 1021 TTTAAAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1058
DB 36192 TTTAAAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 36155

```

```

RESULT 6
ABN67441
ID ABN67441 standard; DNA; 1023 BP.
XX
XX ABN67441;
AC
AC 01-JUL-2002 (first entry)
XX
XX Streptococcus polymnucleotide SEQ ID NO 2795.
DE
DE Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
XX
XX group A streptococcus; Streptococcus pyogenes; antibacterial; gene;
XX
XX anti-inflammatory; infection; vaccine; meningitis; gene therapy; ds.
OS
OS Streptococcus agalactiae.
XX
XX W0200234771-A2.
XX
XX 02-MAY-2002.
XX
XX 29-OCT-2001; 2001MO-GB004789.
XX
XX 27-OCT-2000; 2000GB-00026333.
XX
XX 24-NOV-2000; 2000GB-00028727.
XX
XX 07-MAR-2001; 2001GB-00005640.
XX
XX (CHIR-) CHIRON SPA.
XX
XX (GENO-) INST GENOMIC RES.
XX
XX Telford J, Maignani V, Margarit Y Rost, Grandi G, Fraser C;
XX
XX Tetelin H;
XX
XX WPI; 2002-352536/38.
XX
XX P-PSDB; ABP26810.
XX
XX New Streptococcus protein for the treatment or prevention of infection or
XX
XX disease caused by Streptococcus bacteria, such as meningitis, and for
XX
XX detecting a compound that binds to the protein.
XX
XX Claim 7; Page 3430; 4525pp; English.
XX
XX The invention relates to a protein (ABP25413-ABP30895) from group B
XX
XX streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
XX
XX (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
XX
XX the specification. The proteins have antibacterial and anti-inflammatory
XX
XX activity. (1), nucleic acids encoding (1), ABN6044-ABN71526 and
XX
XX antibodies that bind (1) are used in the manufacture of medicaments for
XX
XX the treatment or prevention of infection or disease caused by
XX
XX Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
XX
XX Nucleic acids encoding (1) are used to detect Streptococcus in a
XX
XX biological sample. (1) is used to determine whether a compound binds to
XX
XX (1). A composition comprising (1) or a nucleic acid encoding (1), may be
XX
XX used as a vaccine or diagnostic composition. The disease caused by
XX
XX Streptococcus that is prevented or treated may be meningitis. Nucleic
XX
XX acid encoding (1) may be used to recombinantly produce (1) and may be
XX
XX used in gene therapy. Antibodies to (1) are used for affinity
XX
XX chromatography, immunoassays, and distinguishing/identifying
XX
XX Streptococcus proteins
XX
XX
SQ
Sequence 1023 BP; 353 A; 162 C; 203 G; 305 T; 0 U; 0 Other;
Query Match 30.3%; Score 321.4; DB 6; Length 1023;
Best Local Similarity 58.8%; Pred. No. 2.1e-58;
Matches 600; Conservative 0; Mismatches 406; Indels 15; Gaps 2;
QY 11 TATATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 70
DB 5 TTTATATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 64
QY 71 TGAATATTAAGGCAAGAGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 130
DB 65 TGGTATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 124

```



```

Db      180 AGAGCAATGCTGTAGAGGCA-----AACACTCATCCAGCAATGAGA 227
Qy      192 AGAATCTGAGTTAGAAATGACAAAAAGCTGAAATGAGTACGATATCTCATCT 251
Db      228 TGTCTCTATTATGAAATTAATGAAACCTTGCCCTGTAATGAAAGATTAATTAAT 287
Qy      252 AATGGAATATGACGCTAGCTTAATTAATTAATGATTTTCTACCCCTTTGAGAGATC 311
Db      288 AAAAGAAATGCTAGAGTAAAGGATATTTTATGAGTACTCTCTCTGCGAGC 347
Qy      312 TATGATTTTATGATCTTTGAAACAAAAATATGAAATCCCTCAGGTGATAT 371
Db      348 TGTCTTACGATTTACAGATATGATATTCACGATATTAATCGGCTTGCGAATGTA 407
Qy      372 GAAATTAACGATCTTTGAAAAAATGCGCAAGCTTCGATCCCTGATTAAGAAATATCAT 431
Db      408 TAACTACCACTTAATTAATGATGAGCTCTT-----TGGTAAAGCTATTAATCT 458
Qy      432 ATCAACAGGATGCTACTATGATGATATAAACAAGCTGCTTCTATTTATTAATA 491
Db      459 CTCTACCGGCAATGATTTCTATTAAGACATCAAAAGTGGTAAATTAATTCAGAGAC 518
Qy      492 TAAAGTTCGGTGTGTAATATATACATATTAATGATGACATCTGAAATCCAGCCCT 551
Db      519 AGGGGTACC-----TTATGCTTGTCTCAGCTACGTAACAATCTAACCAACCCCTTA 569
Qy      552 TGAGATGTAACCTTAATCTATTAATGATTTGAAAAAAGCTTCCCTAAGAAATAC 611
Db      570 CGAAATATGCTGATTTGGTGTATGAAAGATTTATCTAGACCTTTCCAGACGCAATCAT 629
Qy      612 AGGCTTCTGATCATCTGACGAGGCTTTATGACGCTTTCGCGCGGCTTTATGAAAT 671
Db      630 TGGCTCTGCTGACCTACTTATGATTAATGCTTCTGAGAGAGATGATTTAGGCG 689
Qy      672 AACTTTATGAAAAACATTTCATCTTATGATTAATCTATGCTGCGCAATCATTTGCC 731
Db      690 TTGATTTTAAAGCGCTCACTTACTGACCGGATGAGTCCGCAAGTCTGATTTGATG 749
Qy      732 CTCATTAAGACCTGATGAACTGAACATCTTTGATTTGGGCTGAGTGTGTAATAATC 791
Db      750 CTCTATGATTCGGGATCTTTTAAGAGCTCAGAGAGCCCTCATGCTTTAAATTTGCC 809
Qy      752 TTTAGCTTAATTAATGATTAATGCTTACAGCTTCAGAAAGAAATTAATGATGACAG 851
Db      810 ACGCGCGCGCAAAAAAGACACGATATATGCGGAGAAAGCAACTAAAGATTTGCCCTT 869
Qy      852 AAGTCTATTATGCTAAACAGAGATTAATAAAGATGATTTTTCAGAAAAAATAT 911
Db      870 TGCATCTGCTGAGAGATTAAGACATTAATAAAGAGAACTGTTGTCGAGATTAATCT 929
Qy      912 AACCAAAAAAGACCTGCTAATG---TATCAGTCCAGTGAAGTGTATTAATTTATTTGG 968
Db      930 ATGGGTAAAGCCCGAGCAATGAGACTTGCAGCTCAACGAATATGAACATTAATTTGG 989
Qy      969 TAAATTTGAGAGAGAGCTTTATTCAGATTAATTAATTAATTCATCCGATTTCAAAA 1028
Db      990 TAAAGTCTGCTGCTTGCATTAATTCGAAAGGTGCTCAAAATCAAAAATGATTAATTA 1049

```

RESULT 10
ID AAA81502 standard; DNA; 65632 BP.

AAA81502;

04-DEC-2000 (first entry)

N. meningitidis partial DNA sequence gnm_49 SEQ ID NO:49.

Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;
antigen; vaccine; diagnosis; infection; antibacterial; identification;
Meningococcus B; MenB; ds.

```

XX      OS      Neisseria meningitidis.
XX      PN      WC20002430-A2.
XX      PD      20-APR-2000.
XX      PF      08-OCT-1999; 99WC-US023573.
XX      PR      09-OCT-1998; 98US-0103794P.
XX      PR      30-APR-1999; 99US-0132068P.
XX      PA      (CHIR) CHIRON CORP.
XX      PI      Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC,
PI      Meisigant V, Galeotti C, Mora M, Ratti G, Scarfelli M, Scarlato V,
PI      Rappuoli R, Pizza M;
XX      DR      WPI; 2000-318079/27.
XX      PT      Isolated nucleotide sequences of Neisseria meningitidis which can be used
PT      in the diagnosis and treatment of N. meningitidis infection and other
PT      Neisserial infections, for example, N.gonorrhoea.
XX      PS      Claim 7; Page 1331-1350; 1760bp; English.
XX

```

The present invention describes methods of obtaining immunogenic proteins from Neisseria genomic sequences. AA81453 to AA82414 represent specifically claimed Neisseria meningitidis genomic DNA sequences. AA81460 to AA81503 and AA825620 to AA825663 represent Neisseria DNA sequences and their corresponding proteins; AA81254 to AA81259 and AA81304 to AA81321 represent PCR primers used in the isolation of Neisseria meningitidis DNA sequences; and AA81332 to AA81452 represent Neisseria meningitidis MenB polynucleotide ORF sequences, which are all used in the exemplification of the present invention. The nucleic acid sequences, protein sequences, and antibodies against them, can be used in the manufacture of a composition. The composition can be used as a medicament (or in the manufacture of a medicament) for treating, preventing or diagnosing infection due to Neisserial bacteria. For example, some of the identified proteins could be components of vaccines against Meningococcus B; against all serotypes; and/or against all pathogenic Neisseriae. Identification of sequences from the bacterium will also facilitate production of biological probes, particularly organism-specific probes. Attempts to make efficacious Meningococcus B vaccines have failed mainly due to antigen tolerance. Multivalent vaccines have also been tried but none have successfully overcome antigenic variability. The provision of further, complete sequences may provide an opportunity to identify secreted or surface exposed proteins that may be presumed targets for the immune system and which are not antigenically variable or at least more conserved than other more variable regions

Sequence 65632 BP; 16704 A; 16327 C; 15474 G; 17126 T; 0 U; 1 Other;

Query Match 10.2%; Score 107.6; DB 3; Length 65632;

Best Local Similarity 47.4%; Prid. No. 6e-13; Mismatches 504; Indels 33; Gaps 4;

```

Qy      12 ATATATGCTGTGCAAAATTTGGTGAACCAATATGCTAGTGTATTTTCAAGAGAAAT 71
Db      55801 ATTGATTTATCTGTGAATTCGGCATCAATCAATGAAGGCTCTTTAAAAACACCTTTTGAAT 55860
Qy      72 GATTTTAAAGCCAAAGAGCGGCTGTATATGAGTAAATTTCCAAACATTTAAAGCTGA 131
Db      55861 GGTGATGCTGCGATATATGAGCGCTGAAGTTGTTAAACATCAACACACATCTTGA 55920
Qy      132 TAAATTAATTTACGATTAATGACCTAAGGAGAGATCAATTAATAAACAAGAGATTT 191
Db      55921 AGAGAAATGCTGATGAGGCA-----AACAGTCAATTCAGCAATGAGAGA 55968
Qy      192 AGAATCTCAGTTGAATGACAAAAAGCTTGAATGAATTAAGATTAATCTCCACT 251
Db      55969 TGTCTCTATTATTAATTAATGAAAGCTTCCGCGCTGAATGAAGAGATGATTAAT 56028

```

```

QY 252 AATGATATGAGTCAAGTTTAAATTAGATTTTCTACCCCTTGAAGACATC 311
DB 56029 AAAAGATACGTAAGAGTAAAGGTATGATTTTATCAGTACTCTCTCTGTCAGC 56088
QY 312 TATGATTTTAAAGATCTTTGAAACAAAATAATGAAAAATCCCTCAGTGTATTT 371
DB 56089 TGCTTACGATTAACAGTATGATATTTCCAGCATTAATAATGCGCTGCGCAATGTA 56148
QY 372 GAATTTACCGATCTTGAATAAAATAGCCAGCTCCGATCCCTGATTAAGAAATTAATCAT 431
DB 56149 TAACCTACCTTAATTAATTAACCTGCTGCTCTT-----TGTAAAGCTTATATCTT 56199
QY 432 ATCAACAGATGAGTCTACTATGATTAAGATTAACAGTCTGTTCTATTTTAAATAA 491
DB 56200 CTCTACCGGATGATATCTATGTAAGATCAAAAAGTCGTGAAATTTTGAAGAGC 56259
QY 492 TAAAGTCCGCTGTATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 551
DB 56260 AGGGGTAC-----TTATGCTTGTCTCACTGATACCAATCAATCAACCCCTTA 56310
QY 552 TGAAGATGTAACCTTATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 611
DB 56311 CGAAGATGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 56370
QY 612 AGGCTTCTGATCATTTAGACGGTTTAAAGAGCTATTGCGCGGTGCTTATGGAAT 671
DB 56371 TGCGCTGTGACCAATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 56430
QY 672 ACCTTTATGAAAAAATCTTCACTTAAATTAATTAATTAATTAATTAATTAATTAATTAAT 731
DB 56431 TTGATTTTGAAGCTACTTAACTGACCGCATGATGCGCCAGGTGCGATTTGATG 56490
QY 732 CTCAATGAACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 791
DB 56491 CTCTAATGATCCGATCTTAAAGAGCTCAAGCAAGGGCTCAAGCTTAAATTTGCG 56550
QY 792 TTTAGTCAAAATGTAAGTGTACAGCTTCAAGAAAGAAATTAATTAATTAATTAATTAAT 851
DB 56551 ACCGGGGGCAAAAAGACAGATTAACGGGAGAAAACCAATTAATTAATTAATTAATTAAT 56610
QY 852 AAAGTATATATAGTAAACACAGATTAATAAAGAGTGTATTTTCAAAAAATAT 911
DB 56611 TGCATCTGCTGACAGATTAAGACATTAATAAAGAGATCTGTGCGGATTAATCCT 56670
QY 912 AACAAACAAAGACCTGTATG--TATGATCGATGAGTGTATTAATTTGCG 968
DB 56671 ATGGGTAAAGCCCAAGCAATGAGACCTTCAAGCTCAAGCAATTAATTAATTTGCG 56730
QY 969 TAAATGACAGACAGATTTATTCAGATGAATTAATTAATTAATTAATTAATTAATTAATTAAT 1028
DB 56731 TAAAGTCTGCTTGCATATTTGCAAAAGTGTCAATTAATAAATCAATTAATTAATTAAT 56790

```

RESULT 11

AAA81490_00/c
 WP Sequence split into 15 fragments LOCUS AAA81490 Accession Aa81490

Fragment Name	Begin	End
AAA81490_00	1	110000
AAA81490_01	100001	210000
AAA81490_02	200001	310000
AAA81490_03	300001	410000
AAA81490_04	400001	510000
AAA81490_05	500001	610000
AAA81490_06	600001	710000
AAA81490_07	700001	810000
AAA81490_08	800001	910000
AAA81490_09	900001	1010000
AAA81490_10	1000001	1110000
AAA81490_11	1100001	1210000
AAA81490_12	1200001	1310000
AAA81490_13	1300001	1410000
AAA81490_14	1400001	1437668

ID AAA81490 standard; DNA; 1437668 BP.

AC AAA81490;

DT 04-DEC-2000 (first entry)

DE N. meningitidis B full length genome DNA sequence SEQ ID NO:1068.

XX N. meningitidis; Neisseria gonorrhoeae; genome; immunogenic;

XX antigen; vaccine; diagnosis; infection; antibacterial; identification;

XX Meningococcus B; MenB; de.

XX Neisseria meningitidis.

OS Neisseria meningitidis.

PN MO200022430-A2.

XX 20-APR-2000.

PD 20-APR-2000.

PF 08-OCT-1999; 99MO-US023573.

XX 09-OCT-1998; 98US-0103794P.

PR 30-APR-1999; 99US-0132068P.

XX (CHIR) CHIRON CORP.

PA Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC, Masiarant V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V; Rappoli R, Pizza M;

PI MPI; 2000-318079/27.

DR 09-DEC-2000 (first entry).

XX Isolated nucleotide sequences of Neisseria meningitidis which can be used

PT in the diagnosis and treatment of N. meningitidis infection and other

PT Neisseria infections, for example, N.gonorrhoea.

XX Claim 7: Page 866-1272; 1760pp; English.

XX The present invention describes methods of obtaining immunogenic proteins

CC from Neisseria genomic sequences. AAA81493 to AAA82414 represent

CC specifically claimed Neisseria meningitidis genomic DNA sequences;

CC AAA81260 to AAA81303 and AAA825620 to AAA825663 represent Neisseria DNA

CC sequences and their corresponding proteins; AAA81264 to AAA81259 and

CC AAA81304 to AAA81321 represent PCR primers used in the isolation of

CC Neisseria meningitidis DNA sequences; and AAA81322 to AAA81452 represent

CC Neisseria meningitidis MenB polynucleotide ORF sequences, which are all

CC used in the exemplification of the present invention. The nucleic acid

CC sequences, protein sequences, and antibodies against them, can be used in

CC the manufacture of a composition. The composition can be used as a

CC medicament (or in the manufacture of a medicament) for treating,

CC preventing or diagnosing infection due to Neisseria bacteria. For

CC example, some of the identified proteins could be components of vaccines

CC against Meningococcus B; against all serotypes; and/or against all

CC pathogenic Neisseriae. Identification of sequences from the bacterium

CC will also facilitate production of biological probes, particularly

CC organism-specific probes. Attempts to make efficacious Meningococcus B

CC vaccines have failed mainly due to antigen tolerance. Multivalent

CC vaccines have also been tried but none have successfully overcome

CC antigenic variability. The provision of further, complete sequences may

CC provide an opportunity to identify secreted or surface exposed proteins

CC that may be presumed targets for the immune system and which are not

CC antigenically variable or at least more conserved than other more

CC variable regions

XX Sequence 1437668 BP; 344338A; 353206C; 385074G; 355045T; 0U; 50cher;

Query Match 10.2%; Score 107.6; DB 3; Length 110000;

Best Local Similarity 47.4%; Pred. No. 6.5e-13;

Matches 483; Conservative 0; Mismatches 504; Indels 33; Gaps 4;

QY 12 ATATATCGTGTGTAATTTGTTGCAACCATATAGTGTGTTATATGCAAGAAAT 71

DB 78357 ATGATATCTGTGAATTCGGATCATATCAAGAGCTTTAAACAGCTTTGAAAT 78298

```

QY 72 GATATTAAAGCCAGAGCCGGGTGTAATGCAAGTAAATCCAAATTTAAAGCTGA 131
DB 78297 GGTTCATGCTGCTTAAATGACGGCGCTGAAGTTTAAACCTCAAAACACATCTTGA 78238
QY 132 TAAATTAATTCAGCTATGCACTTAAGCCAGATATCAATTAATAAAACACAGGAATT 191
DB 78237 AGACGAATGCTGATGAGGCCA-----ACAGATCTCCAGGCAATGACGA 78190
QY 192 AGAATCTAGTAAATGACAAATAAGCTGAATGAAGTATGAGATATTCCTCACT 251
DB 78189 TGTCTCTATTATTAATTAATGAAAGCTTGCCTCGATGAAGAAAGATTAATT 78130
QY 252 AATGGAATATGCAATGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 311
DB 78129 AAAAATAATGATAGAGATGATGATGATTTTATCAAGTATCTCTCTCTGCGAGC 78070
QY 312 TATTGATTTTATGATCTTTGAAACAAATAATGAAATCCCTTCAGGTGATATT 371
DB 78069 TGCCTTACGATTAACAAGTATGATATTCAGCATATTAATAATCGGCTCTGGGAATGTA 78010
QY 372 GAATTTACCGGTATCTGAAATAATAGCCAGCTCCAGTCCCTGATTAAGAAATATCAT 431
DB 78009 TAACATACCATTAATTAATCTGCTGCTCTT-----TGGTAAAGCTATTAATCT 77959
QY 432 ATCAAACGGAATGCTACTATTTATGATGATGATTAATAAGCTCTGTTCTATTATTAATPA 491
DB 77958 CTCTACCGGATGAAATTCATTAATGAAAGCATCAAAAGCTGATGAATAATATTCGGAAGC 77899
QY 492 TAAAGTTCGGTGTGAATATACAAATTAATTAATTAATTAATTAATTAATTAATTAAT 551
DB 77898 AAGGGTACC-----TTAGCTTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 77848
QY 552 TGAGATATGTAACCTTAATGCTATTAATGATTTGAAAAACATTCCTTAAGATTAATCAT 611
DB 77847 CGAAGATGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 77788
QY 612 AAGCTTCTGATGATCTGAGCGGTTTATGACGATGATGATGATGATGATGATGATGATGAT 671
DB 77787 TGGCTCTGTCACCAATCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 77728
QY 672 AACTTTATGAAAAACATTTCACTTATGATTAATTAATTAATTAATTAATTAATTAATTAAT 731
DB 77727 TTGCAATTTAGAGCGTCACTTATGACGCGATGATGATGATGATGATGATGATGATGAT 77668
QY 732 CTCAATAGACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 791
DB 77667 CTCTATGATTCGGAATCTTTTAAAGCTCAAGCAAGCGCTCATGCTTTAAATTTGCT 77608
QY 792 TTTAGCTTCAATAGTAAAGTGTATGACGCTCAGAAAGGATTAATTAATTAATTAATTAAT 851
DB 77607 AAGCGCGGCAAAAGACATTAATGCGGGAAGAAAGCACTTAAGATTTGCTCT 77548
QY 852 AAGTCTATTAATGCTAAACAGATTAATAAAGGTGAGTTTTCAGAAAAATTAAT 911
DB 77547 TGCATCTGCTGATGAGATTAAGATTAATAAAGGAGAAAGCTGTCGCGAGATTAATCT 77488
QY 912 AACACAAAAGACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 968
DB 77487 ATGGGTAAAGCCCGGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 77428
QY 969 TAAATTTGAGAGCAAGCTTATTCAGATGATTAATTAATTAATTAATTAATTAATTAATTAAT 1028
DB 77427 TAAAGTCTGCTGCTGCAATTAATTCGAAAGGTGCTCAATTAATAAAGCTGATTAATTAAT 77368

```

RESULT 12
AAFP21544/C
ID AAFP21544 standard; DNA, 349980 BP.

AC AAFP21544;
XX
DT 13-MAR-2001 (first entry)
XX

```

DE Neisseria meningitidis B nucleotide sequence SEQ ID NO:1.
XX
XX Neisseria meningitidis, Neisseria gonorrhoeae, immunogenic; vaccine;
KM diagnosis; antigen; detection; infection; gene therapy; antibacterial;
XX ds.
XX Neisseria meningitidis.
OS
XX W0200066791-AL.
XX
XX 09-NOV-2000.
XX
XX 08-MAR-2000; 2000MO-US005928.
XX
XX 30-APR-1999; 99US-0132068P.
XX 08-OCT-1999; 99MO-US023573.
XX 28-FEB-2000; 2000GB-00004695.
XX
XX (CHIR ) CHIRON CORP.
XX (GENO-) INST GENOMIC RES.
XX
XX Piza M, Hickey E, Peterson J, Tettelin H, Venter JC,
XX Maignani V, Galeotti C, Mora M, Ratti G, Scarcelli M, Scarlato V,
XX Rappuoli R, Frazer CM, Grandi G;
XX WPI; 2000-647603/62.
XX
XX Neisseria meningitidis B full length genome sequence and open reading
XX frames are used to detect, treat and prevent Neisserial infections.
XX
XX Claim 7; Appendix A; 692pp; English.
XX
XX The present invention describes the full length genome of Neisseria
XX meningitidis B (NMB). The sequences in AAFP21544 and AAFP21607 to AAFP21613
XX represent fragments of the NMB genomic sequence, and the sequence was too
XX long to go in a record on its own it was split into 8 sequences which
XX overlap each other at the beginning and end of each sequence by 49980 bp
XX (i.e. the last 49980 bp of AAFP21544 is repeated at the beginning of
XX AAFP21607, the last 49980 bp of AAFP21607 are repeated at the beginning of
XX AAFP21608, and so on). AAFP21545 to AAFP21588 encode the Neisseria proteins
XX given in AAB85530 to AAB85593, and AAFP21589 to AAFP21606 represent PCR
XX primers which are used in the exemplification of the present invention.
XX The NMB genome and fragments from it have antibacterial activity, and can
XX be used in vaccines and gene therapy. Neisseria nucleic acids, proteins
XX and/or antibodies which binds to the proteins can be used in compositions
XX for treating or preventing infection due to Neisserial bacteria or as a
XX diagnostic reagent for detecting the presence of Neisserial bacteria or
XX of antibodies raised to Neisserial bacteria. Computers, computer memory,
XX computer storage medium or computer databases can be used in a search to
XX identify open reading frames (ORFs) or coding sequences within the NMB
XX genome. The DNA sequences provide further opportunities to find antigenic
XX or immunogenic proteins which are more effective in vaccines than the
XX outer membrane proteins currently used
XX
XX Sequence 349980 BP; 83241 A; 85091 C; 95206 G; 86442 T; 0 U; 0 Other;
XX
XX Query Match 10.2%; Score 107.6; DB 3; Length 349980;
XX Best Local Similarity 47.4%; Pred. No. 8e-13;
XX Matches 483; Conservative 0; Mismatches 504; Indels 33; Gaps 4;

```

```

Dh 78189 TGTCTTATTATGAAATTTATGAGACGTGGCCCTGAAAGAAAGATGATTAAT 78130
Qy 252 AATGGAATATGACGTCAGTTTAAATTTAGATTTTTCACCCCTTTTGAAGAACTC 311
Dh 78129 AAAAGATACGATGAGAGTAAAGGTATGATTTTATGAGTACCTTCTCTGTGAGC 78070
Qy 312 TATGATTTTATGACATCTTTGAAACAAAATATGAAAAATCCCTCAGGTGATAT 371
Dh 78069 TGTCTTACGATTAACGATATGATATCCAGCATATTAATATGGCTCTGCGAATTA 78010
Qy 372 GAATTTACCGTATCTTGAATAAATATGCGAACCTTCCATCCCTGATAGAAATATCAT 431
Dh 78009 TAATACCCATTAATTAATTAACGTGGCTCTTCTTCTTCTTCTTCTTCTTCTTCT 77959
Qy 432 ATCAACGGAATGCTCTATATGATGATTAACGATCTTCTTCTTCTTCTTCTTCTTCTT 491
Dh 77958 CTCTACCGGATGATATCTTATGAAAGCATGAAAGCTGATGAAATTTATCGAGAGC 77899
Qy 492 TAAAGTCCGCTTGATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 551
Dh 77898 AGGGGTACC-----TATGCTTGTCTCACTGATACCAACATCTACCAACCCCTTA 77848
Qy 552 TGAGATGATTAACCTTAATGCTATTAATGATTTGAAAAACACTCCCTTAAGATTAATCAT 611
Dh 77847 CGAAGATGTTGATGGGTGATGAAAGATTTATGAAAGCTTTCAGAGCCATCAT 77788
Qy 612 AGGCTCTCTGATCATCTTACGCGGTTTATGACGCTATTCGCGGCTTCTTATGAAAT 671
Dh 77787 TGGCTGTCTGACCAATCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 77728
Qy 672 AACTTTATGAAAAAATTTCACTTAATTAATTAATTAATTAATTAATTAATTAATTA 77731
Dh 77727 TTCCATTTTATGAGGCTCACTTACTGACCGCATGATCGCCAGGCTCGATATTAATG 77668
Qy 732 CTCAATGAAACCTGATGACGAAACATCTTATGATGGGTGAGGTGTTGAAAAATC 791
Dh 77667 CTCTATGATCGGATCTTATTAAGACTTAAGCAAGGCGCTCATGCTTTAAATTTGGC 77608
Qy 792 TTTAGTTCATTAATGATTAAGGTGATTAAGCTTCAAGAAAGAAATTAATTAATTAAT 851
Dh 77607 ACGGCGCGGCAAAAGCAAGATTAATCGCGGAAAGCACTAAAGATTTGCGCTT 77548
Qy 852 AAGCTCTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 911
Dh 77547 TGCATCTGTCTGATGATTAAGCACTTAATTAATTAATTAATTAATTAATTAATTA 77488
Qy 912 AACAAACAAAAGACCTGATTAATG---TATCATGTCGATGAGTATTAATTAATTTGCG 968
Dh 77487 ATGGGTTAAAGCCGAGCAATGAGACTTCAAGCTCAAGAAATTAATTAATTAATTTGG 77428
Qy 969 TAAATTTGACAGAGCAATTTATTAATTAATTAATTAATTAATTAATTAATTAATTA 1028
Dh 77427 TAAAGTGTCTGCTTGCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 77368

```

RESULT 13

AAAS3720 standard; DNA; 11474 BP.

AAAS3720;
 AC AAA53720;
 XX 15-SEP-2003 (revised)
 DT 22-DEC-2000 (first entry)
 XX

DE LPS core biosynthesis locus.

XX Biosynthetic locus; biosynthesis; lipid A biosynthesis;
 XX acetyltransferase; glycosyltransferase; Beta-1,4-GalNAc transferase;
 KW Beta-1,3-galactosyltransferase; alpha-2,3-sialyltransferase;
 KW sialic acid synthase; CMP-sialic acid synthetase; mimic; antibody;
 KW immunity; immunogen; ganglioside; ds.
 XX

```

OS Campylobacter jejuni, O4:384.
FH Location/Qualifiers
FT CDS
FT /tag= a
FT /product= "acyltransferase"
FT /note= "Open reading frame 2b"
FT CDS
FT /tag= b
FT /product= "glycosyl transferase"
FT /note= "Open reading frame 3a"
FT CDS
FT /tag= c
FT /product= "glycosyl transferase"
FT /note= "Open reading frame 4b"
FT CDS
FT /tag= d
FT /product= "beta 1,4-GalNAc transferase"
FT /note= "Open reading frame 5a"
FT CDS
FT /tag= e
FT /product= "beta 1,2- galactosyltransferase"
FT /note= "Open reading frame 6a"
FT CDS
FT /tag= f
FT /product= "GstII sialyltransferase"
FT /note= "Open reading frame 7a"
FT CDS
FT /tag= g
FT /product= "sialic acid synthase"
FT /note= "Open reading frame 8a"
FT CDS
FT /tag= h
FT /product= "sialic acid biosynthetic enzyme"
FT /note= "Open reading frame 9a"
FT CDS
FT /tag= i
FT /product= "CMP-sialic acid synthetase"
FT /note= "Open reading frame 10a"
FT CDS
FT /tag= j
FT /product= "lipooligosaccharide biosynthetic enzyme"
FT /note= "Open reading frame 12a"

```

W0200046379-A1.
 10-AUG-2000.
 01-FEB-2000; 2000MO-CA000086.
 01-FEB-1999; 98US-0118213P.
 31-JUN-2000; 2000US-00495406.
 (CANADA) NAT RES COUNCIL CANADA.
 Gilbert M, Wakarchuk WM;
 WPI: 2000-524418/47.
 DR P-PSDB, AA97200, AA97201, AA97202, AA97203, AA97204, AA97205,
 DR AA97206.
 Novel glycosyltransferase polypeptides and polynucleotides useful for
 biosynthesis of ganglioside and ganglioside mimics, as diagnostic
 reagents and as immunogen for producing antibodies.
 Claim 1; Page 86-90; 120pp; English.
 A reaction mixture for the synthesis of a sialylated oligosaccharide is
 useful for synthesizing sialylated oligosaccharides such as ganglioside,
 lysoganglioside or their mimics. Glycosyltransferases are useful for
 chemo-enzymatic synthesis of oligosaccharides, including gangliosides and
 other oligosaccharides that have biological activity. The enzymes and
 nucleic acids that encode them are useful for studies of the pathogenesis

mechanisms of organisms that synthesize ganglioside mimics, such as C. jejuni and the nucleic acids are used as probes to study expression of genes involved in ganglioside mimetic synthesis. Antibodies raised against the glycosyltransferases are also useful for analyzing the expression patterns of these genes involved in pathogenesis. The nucleic acids are also useful for designing antisense oligonucleotides for inhibiting expression of the Campylobacter enzymes that are involved in the biosynthesis of ganglioside mimics that can mask the pathogens from the host's immune system. The oligosaccharides are useful as diagnosing reagents or as therapeutics and as immunogens for producing antibodies. Bacterial glycosyltransferase can be used to catalyze the formation of oligosaccharides that are identical to the corresponding mammalian structures and are easier and less expensive to produce in large quantity, compared to the mammalian glycosyltransferase. The bacterial origin of the enzymes facilitates expression of large quantities of the enzymes using relatively inexpensive prokaryotic expression systems. (Updated on 15-SEP-2003 to standardise OS field)

Sequence 11474 BP; 4506 A; 1495 C; 1613 G; 3860 T; 0 U; 0 Other;

Query Match 9.9%; Score 105.2; DB 3; Length 11474;
Best Local Similarity 49.1%; Pred. No. 1.4e-12;
Matches 479; Conservative 0; Mismatches 458; Indels 39; Gaps 6;

17 TCGTCTGCTGAATTTGCTGCACCAATTAAGTAGTGTATTTGCAAGAGAAATGATAT 76
DB TCGTCTGCTGAATTAAGCAATTAATCAATTAAGTAGTGTATTTGCAAGAGAAATGATAT 7038
QY 77 TAAAGCCAAAGAGCGCGGTGTTAATGACGTAATTAATCCAAATTTAAAGCTGTAAT 136
DB ATGACACCTTTAGCAGAGGCTTAAGATTAATTAAGCAATTAATCCAAATTTGTAAGAG 7098
QY 137 TAATTCAGCTATGCACTTAAGGCGAGATTTCAATTAATTAATTAATTAATTAATTAAT 196
DB 7099 AGATGAG-----TAAAGCGCTAATAAATAATTTCCGTGTAATGCAAAATTA 7146
QY 197 CTCAGTTAGAAATGACAAATAAGCTGAATGAAGATGACATTAATCTCCATTAATAG 256
DB 7147 GCATTATGATATGCAAAATATGCTTTAGATTAATTAATTAATTAATTAATTAATTAAG 7206
QY 257 AATATGAGTCACTTAATTAATTAATTAATTTTCTTACCCCTTTTGAAGAGCTATATG 316
DB 7207 AATATGAGTCACTTAATTAATTAATTAATTTTCTTACCCCTTTTGAAGAGCTATATG 7266
QY 317 AATTTTATGATCTTTGAATAAATAATTAATTAATTAATTAATTAATTAATTAATTAAT 376
DB 7267 ACCGCTTAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 7326
QY 377 TACCGTATCTTGAATAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 436
DB 7327 ATCCGCTTATTAACACATAG-----CAGCCTTTAAATAAGCTTAATGAATGATGAGA 7377
QY 437 CAGGATGCTCTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAG 496
DB 7378 CAGGATGCTCTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAG 7437
QY 497 TTCCGGTGGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 556
DB 7438 TTCCGGTGGT-----TTAATGACACAAACATCTTAATTAATTAATTAATTAATTAAT 7488
QY 557 ATGTAACCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAG 616
DB 7489 TTGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAG 7545
QY 617 TCTCTGATCATTTACCGGCTTTTATGACGCTAATTCGCGGCTGCTTAATTAATTAATTAAG 676
DB 7546 TAAGCAGACACAAACAGATAATCTTGCGTTAGTACGCTGCTGCTGCTGCTGCTGCTG 7605
QY 677 TTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAG 736
DB 7606 TCGTTGAAGACATTTTACTGATGATGATGATGATGATGATGATGATGATGATGATGATG 7665
QY 737 TAGAAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 796

DB 7666 TGGATACACAGCGCTTTTAAATAAGCTTATTAATTAATTAATTAATTAATTAATTAAG 7725
QY 797 GTTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAG 856
DB 7726 GATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAG 7785
QY 857 CTATTATGCTTAAAC--AGAGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 913
DB 7786 CAAGCGTGTGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 7845
QY 914 CAACAAAGACCTG--TAATGCTATCAGTCCGATGAGTGGTATTAATTAATTAATTAAG 970
DB 7846 CGGTTAAAGACCTGAGCTTGTGATTAATTAATTAATTAATTAATTAATTAATTAATTAAG 7905
QY 971 AAATTGACAGACAGA 986
DB 7906 AAAAGCATTAAGAGA 7921

RESULT 14

AB13665 standard; DNA; 11474 BP.

AB13665;

07-FEB-2003 (first entry)

Campylobacter jejuni genomic lipooligosaccharide biosynthesis locus.

Gene; ds; enzyme; gene therapy; acyltransferase; glycosyltransferase; GalNAc transferase; N-Acetylglucosamine transferase;

galactosyltransferase; sialyltransferase; sialic acid synthase;

cytidine 5'-monophosphate sialic acid synthetase;

CMP sialic acid synthetase; acetyltransferase; ganglioside synthesis;

ganglioside mimetics; inflammation; tumour metastasis.

Campylobacter jejuni.

W0200274942-A2.

26-SEP-2002.

22-FEB-2002; 2002WO-CA000229.

21-MAR-2001; 2001US-00816028.

(CANADA) NAT RES COUNCIL CANADA.

Gilbert M, Wakarchuk MW;

WPI; 2003-040554/03.

Claim 1; Page 89-95; 107pp; English.

New glycosyltransferases from Campylobacter, useful for synthesizing gangliosides and ganglioside mimetics, and in studying the pathogenesis mechanisms of organisms that synthesize ganglioside mimetics.

The invention comprises the amino acid and coding sequences of Campylobacter jejuni proteins. The C. jejuni proteins of the invention may be either an acyltransferase; glycosyltransferase; GalNAc (N-Acetylglucosamine) transferase; galactosyltransferase;

sialyltransferase; sialic acid synthase; cytidine 5'-monophosphate (CMP) sialic acid synthetase; acetyltransferase. The C. jejuni DNA and protein sequences of the invention are useful for ganglioside synthesis, studying ganglioside mimetics, and for designing oligonucleotides to inhibit expression of Campylobacter enzymes involved in the biosynthesis of ganglioside mimetics that can mask the pathogen's from the host's immune system. The C. jejuni oligosaccharides of the invention may be used as diagnostic reagents (e.g. to locate areas of inflammation or tumour metastasis). The present DNA sequence represents a Campylobacter jejuni gene of the invention

XX Sequence 11474 BP; 4506 A; 1495 C; 1613 G; 3860 T; 0 U; 0 Other;
 SQ Query Match 9.9%; Score 105.2; DB 7; Length 11474;
 Best Local Similarity 49.1%; Pred. No. 1.4e-12;
 Matches 479; Conservative 0; Mismatches 458; Indels 39; Gaps 6;

17 TCGTTCCTGAATGCTTGGCAACCATTAATGCTGTTGATGATGCAAGAGAAATGATAT 76
 DB TCGTTCCTGAATGCTTGGCAATTAATGCTGTTGATGATGCAAGAGAAATGATAT 7038
 6979 TCGTTCCTGAATGCTTGGCAATTAATGCTGTTGATGATGCAAGAGAAATGATAT 7038
 77 TAAAGCCAAAGGCGCGTGTATGATGCTGTTGATGATGCAAGAGAAATGATAT 136
 DB ATGCAAGCTTTAGCAAGGCTGCTGATGATGATGATGCAAGAGAAATGATAT 7098
 7039 ATGCAAGCTTTAGCAAGGCTGCTGATGATGATGATGCAAGAGAAATGATAT 7098
 137 TAAATTCAGCTATGCACTTAAAGCAAGATGATGATGATGCAAGAGAAATGATAT 196
 DB AGATGAG-----TAAAGCGCTTAAAGAAATGCTGCTGATGCAAGAAATTA 7146
 7099 AGATGAG-----TAAAGCGCTTAAAGAAATGCTGCTGATGCAAGAAATTA 7146
 197 CTCAGTTAGAAATGCAAAAGCTGATGATGATGATGATGCAAGAGAAATGATAT 256
 DB GCATTTAGATGATGCAAAAGCTGATGATGATGATGATGCAAGAGAAATGATAT 7206
 7147 GCATTTAGATGATGCAAAAGCTGATGATGATGATGATGCAAGAGAAATGATAT 7206
 257 AATATGCAAGCTGATGATGATGATGATGATGATGATGCAAGAGAAATGATAT 316
 DB AATATGCAAGAAATTAAGCTGATGATGATGATGATGATGCAAGAGAAATGATAT 7266
 7207 AATATGCAAGAAATTAAGCTGATGATGATGATGATGATGCAAGAGAAATGATAT 7266
 317 ATTTTATGATGATGATGATGATGATGATGATGATGATGCAAGAGAAATGATAT 376
 DB ACCGCTTAAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 7326
 7267 ACCGCTTAAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 7326
 377 TACCGTATCTGAAAGAAATGCAAGCTTCCATGCTGATGATGATGATGATGAT 436
 DB ATCCGCTATTAAGCAATG-----CACCTTTAAAGAAATGATGATGATGAT 7377
 7327 ATCCGCTATTAAGCAATG-----CACCTTTAAAGAAATGATGATGATGAT 7377
 437 CAGGAATGCTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 496
 DB CAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 7437
 7378 CAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 7437
 497 TCCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 556
 DB TCCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 7438
 7438 TCCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 7438
 557 ATGTAACCTTAATGCTATGATGATGATGATGATGATGATGATGATGATGAT 616
 DB TTTAAGATTAAGCTATGCTGATGATGATGATGATGATGATGATGATGATGAT 7489
 7489 TTTAAGATTAAGCTATGCTGATGATGATGATGATGATGATGATGATGATGAT 7489
 617 TCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 676
 DB TTAGCAACCAACAGATGATGATGATGATGATGATGATGATGATGATGATGAT 7546
 7546 TTAGCAACCAACAGATGATGATGATGATGATGATGATGATGATGATGATGAT 7546
 677 TTTAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 736
 DB TCGTTAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 7606
 7606 TCGTTAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 7606
 737 TAAAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 796
 DB TGAATTAAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 7725
 7666 TGAATTAAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 7725
 797 GTTCAATTAAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 856
 DB GAAATTAAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 7785
 7726 GAAATTAAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 7785
 857 CTATTAAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 913
 DB CAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 7845
 7786 CAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 7845
 914 CAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 970
 DB GGGTAAAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 7905
 7846 GGGTAAAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 7905
 971 AAATGCAAGCAAGA 986

DB 7906 AAAAGCAATTAAGCA 7921

RESULT 15
 ABK37785
 ID ABK37785 standard; DNA; 1000 BP.
 XX ABK37785;
 AC
 XX
 XX
 DT 08-MAY-2002 (first entry)
 XX
 XX
 DE DNA sequence upstream of slad gene.
 XX
 XX
 KW Upstream sequence; ds; Antibacterial; vaccine; bleb;
 KW Gram-negative bacteria; outer membrane; LPS; lipopolysaccharide;
 KW meningitis; bacteraemia; otitis media; pneumonia; chronic bronchitis;
 KW sinusitis.
 XX
 XX
 OS Neisseria meningitidis serogroup B.
 XX
 XX
 PN MO200209746-A2.
 PD 07-FEB-2002.
 XX
 XX
 PF 31-JUL-2001; 2001MO-EP008857.
 XX
 XX
 PR 31-JUL-2000; 2000MO-EP007424.
 PR 08-FEB-2001; 2001GB-00003170.
 XX
 XX
 PA (SMK) SKITHLINE BECHAM BIOLOGICALS.
 XX
 PI Berthel FJ, Dalemans W, Dencel P, Deguesne G, Feron C, Garcon N,
 PI Lobet Y, Poolman J, Thiry G, Thonard J, Voet P,
 XX MPI, 2002-188688/24.
 XX
 PT New immunogenic composition comprising an antigen derived from a pathogen
 PT and a blep preparation from Neisseria meningitidis, useful as a vaccine
 PT for treating or preventing disease caused by the pathogen.
 XX
 PS Disclosure; Page 87; 125pp; English.

XX The invention relates to an immunogenic composition comprising an antigen
 CC derived from a pathogen capable of protecting a host against the
 CC pathogen, mixed with an adjuvant comprising a bleb preparation derived
 CC from a Gram-negative bacterial strain. The immunogenic composition
 CC consists of N. meningitidis B blebs or N. meningitidis C polysaccharide
 CC antigen. The blebs (derived from the outer membrane) may also have their
 CC toxic lipopolysaccharide (LPS) content reduced using heterologous down
 CC regulating sequences for LPS pathway genes or by up regulating genes
 CC involved in LPS synthesis suppression. By a promoter replacement
 CC technique. The immunogenic preparation is useful in the manufacture of a
 CC medicament for the treatment of a disease caused by the pathogen from
 CC which the antigen is derived (e.g. from Neisseria, meningitis and
 CC bacteraemia, from Moraxella, otitis media and pneumonia, and from H.
 CC influenzae chronic bronchitis, sinusitis, pneumonia and otitis media).
 CC The bleb derived from M. catarrhalis or from a non-typable H. influenzae
 CC is useful as an adjuvant in an immunogenic composition comprising one or
 CC more pneumococcal capsular polysaccharides or protein antigens. The
 CC present sequence is an upstream sequence from an N. meningitidis,
 CC H. influenzae or M. catarrhalis gene involved in LPS biosynthesis, which
 CC either up regulates or down regulates sequences to which it is attached
 CC
 XX

SO Sequence 1000 BP; 315 A; 192 C; 208 G; 285 T; 0 U; 0 Other;
 SQ Query Match 9.4%; Score 100; DB 6; Length 1000;
 Best Local Similarity 47.8%; Pred. No. 1.2e-11;
 Matches 400; Conservative 0; Mismatches 415; Indels 21; Gaps 3;

106 TCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 255
 DB TCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 158

Thu May 13 11:53:11 2004

us-09-930-440b-7.rng

Page 16

```
QY 256 GAATATGCGATCGATTAAATTAGATGTTTTTCTACCCCTTTGAGAGAGACTTAT 315
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 159 GAATACGTAGAGAGTAAAGGATATGATTTTATACAGTACTCCCTTCTCTGCGACGTGCT 218
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 316 GATTTTATAGCATCTTTGAAACAAAATAATGAAATCCCTTCAGGTAGTTATGAT 375
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 219 TTACATATACACAGTATGATATTCACATATATAATCGGCTCGGCAATGTATTAAC 278
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 376 TTACCGTATCTTGAATAATACCAAGCTTCGATCCCTGATAGAAATATCATCA 435
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 279 TACCATTAATTAACCTGGTGGCCCTCTT-----TGGTAAAGCTTATATCTCTCT 329
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 436 ACAGGAATGGCTACTATGATAGATAAACAGCTGTTTCTATTTTATATAATATAA 495
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 330 ACCGGCATGAATCTATTGAAAGACATCAAAAAGTCGTAGAAATTTTCGAGAGCAGG 389
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 496 GTTCGGTGTGTAAATATACATATATACATTGCAATGCAATCCAGCCCTTTAG 555
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 390 GTACCTTATGCTTGTCT-----CACTGACCAACATCTACCCCAACCCCTTACGAA 440
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 556 GATGTAAACCTTAATGCTATTTATATGATTTGAAAAACCTTCCCTAAGATATACATAGG 615
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 441 GATGTTCGATGGGTAGTATGACATTTATCTGAAGCTTTTCAGAGCAATCATGSC 500
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 616 TTCTCTGATCATTTCTAGCGGGTTTATGCAAGCTATTCGCGCGTGCCTTANGAATTA 675
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 501 CTGCTCGACCATACCTTATAGTACTATGCTTGTAGAGCAGTAGCTTAGCGGCTCG 560
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 676 TTTATGAAAAACCTTCACTTATAGTAAATCTATGCTGCGCCAGATCATTTGGCCTCA 735
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 561 ATTTTAGAGCGTCACTTACTAGACGCAATGATCGCCAGTCCGATATGTATGCTCT 620
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 736 ATAGAACCTGATGAACTGAACATCTTGTATGCGGTGCGGTGCTGTAATAATCTT 795
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 621 ATGAATCCGATACCTTTTAAAGGCTCAGCAAGGGCTCATGCTTAAAAATTGGCAGGC 680
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 796 GGTTCAAATAGTAAAGTGTACAGCTTACAGAAAGAAATTAATCGTACCAAGAAAG 855
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 681 GGCGGCAAAAAGACACGATTAATCGCGGAGAAAGCCCACTAAAGATTTGCCCTTGCA 740
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 856 TCTATATAGCTTAAACAGAGATTAATAAAGGTGAGGTTTTTTCAGAAAAAATATAACA 915
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 741 TCTGTCTAGCAGATTAAGACATTAATAAAGAGAACTGTCTCCGAGATAACCTATGG 800
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 916 ACAAAAAGACCTGTATAGG---TATCAGTCCGATGAGGTGTATATTATTTGGTTAA 972
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 801 GTTAAACGCCCAAGGCAATGAGACTTCAGGTCACGAATATGAAACATTATTTGGTTAG 860
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 973 ATTGCAGAGCAAGCTTATTCAGATGAATTAATATCATAGCGAATTCAAAAA 1028
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 861 GTGCTGTGCTGCAATATTCGCAAGGTCCTCAATCAAAAAAAGTATATGAAAT 916
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

Search completed: May 10, 2004, 08:08:15
Job time : 464.589 secs

Thu May 13 11:53:12 2004

us-09-930-440B-7.rst

Page 1

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2004, 07:15:33 : Search time 3062.04 Seconds
(Without alignments)
10327.779 Million cell updates/sec

Title: US-09-930-440B-7
Perfect score: 1059
Sequence: 1 atgaggaacatataatgcgt.....aattgagacacaaatattatg 1059

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database:

EST:
1: em_estda:
2: em_estum:
3: em_estin:
4: em_estnu:
5: em_estov:
6: em_estpl:
7: em_estro:
8: em_hic:
9: gb_est1:
10: gb_est2:
11: gb_hic:
12: gb_est3:
13: gb_est4:
14: gb_est5:
15: em_estfun:
16: em_estom:
17: em_gsa_hum:
18: em_gsa_hiv:
19: em_gsa_pln:
20: em_gsa_vic:
21: em_gsa_fun:
22: em_gsa_mam:
23: em_gsa_mus:
24: em_gsa_pro:
25: em_gsa_rtd:
26: em_gsa_pbg:
27: em_gsa_vir:
28: gb_gsa1:
29: gb_gsa2:

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	88	8.3	649	14	CF998654 AGENCOURT
2	79.4	7.5	1200	13	EX437758 EX437758
3	78.8	7.4	1435	11	BC050955 Danto rer
4	73.8	7.0	876	14	CD327087 AGENCOURT

5	71.8	6.8	508	14	CA550431
6	69.4	6.6	545	12	B1377301
7	68.6	6.5	715	12	B185296
8	68.6	6.5	868	14	CD301354
9	68.2	6.4	676	12	BM095216
10	67.6	6.3	673	14	CF360452
11	67.6	6.3	1225	29	CNS0161D
12	66.4	6.3	613	13	BQ600407
13	66.2	6.3	643	12	B069790
14	66.2	6.3	678	14	CB530032
15	65.6	6.2	469	14	CB73943
16	65.6	6.2	1070	9	AL54411
17	65.2	6.2	417	10	AM416754
18	65.2	6.2	886	28	BH177277
19	65.2	6.2	886	29	CNS07JUX
20	65.2	6.2	1101	29	CNS0039G
21	65	6.1	1201	13	BX335216
22	64.8	6.1	574	14	CA417649
23	64.6	6.1	590	12	BQ014133
24	64.6	6.1	710	10	BR448706
25	64.6	6.1	777	12	BG568099
26	64.2	6.1	576	14	CD365322
27	63.6	6.0	579	14	CA314981
28	63.4	6.0	653	14	CB427711
29	63	5.9	551	12	B6179516
30	63	5.9	569	10	AM249532
31	63	5.9	576	12	BM792591
32	63	5.9	577	13	BU626163
33	63	5.9	577	13	CA429806
34	63	5.9	579	13	BQ773353
35	63	5.9	604	14	CA434954
36	63	5.9	608	14	CD671274
37	63	5.9	610	12	BQ001230
38	63	5.9	613	13	BU676733
39	63	5.9	677	14	CA411576
40	63	5.9	678	13	BU687286
41	63	5.9	689	12	BM976474
42	63	5.9	694	13	BQ446501
43	63	5.9	698	12	BM982564
44	63	5.9	769	13	BU618375
45	63	5.9	779	14	CA417480

ALIGNMENTS

CA550431 C0830ED06-
B1377301 BPLG3_001
B185296 UNL-P-FN-
CD301354 AGENCOURT
BM095216 fv31g11.x
CF360452 821784 MA
AL106171 Drosophila
BQ600407 MI-P-E7-a
B069790 B1069790
CB530032 UI-H-FT2-
CB73943 AMGNMCM
AL54411 AL54411
AM416754 52368 MAR
BH177277 008_L_22-
AL614235 T3 end of
AL063921 Drosophila
BX335216 BX335216
CA417649 UI-H-FE0-
BQ014133 UI-H-ED1-
BR448706 7n93d07.x
BG568099 602586980
CD365322 UI-H-FT2-
CA314981 UI-H-FE0-
CB427711 603306 MA
B6179516 602328112
AM249532 2821389.3
BM792591 K-EST0072
BU626163 UI-H-FG1-
CA429806 UI-H-FH1-
BQ773353 UI-H-FE0-
BO773353 UI-H-FE0-
CA434954 UI-H-DH0-
CD671274 fg02906.x
BQ001230 UI-H-DH1-
BU676733 UI-H-ED1-
CA411576 UI-H-ED1-
BU687286 UI-H-ED1-
BM976474 UI-H-ED1-
BQ446501 UI-H-ED1-
BM982564 UI-H-ED1-
BU618375 UI-H-FH1-
CA417480 UI-H-FE0-

RESULT 1
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLES
JOURNAL
COMMENT

CF998654 649 bp mRNA linear EST 25-NOV-2003
AGENCOURT_16388159 NIH_ZGC_7 Danto rerio cDNA clone IMAGE:7039969
5' mRNA sequence.
CF998654
CF998654.1 GI:38519505
EST
Danto rerio (zebrafish)
Danto rerio
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
1 (bases 1 to 649)
NIH-MGC <http://mgc.ncl.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Daniela S. Gerhard, Ph.D.
Office of Cancer Genomics
National Cancer Institute / NIH
Bldg. 31 Rm10A07 Bethesda, MD 20892
Email: cgaps-remail.nih.gov
Tissue Procurement: Len Zou, Harvard
cDNA Library Preparation: Open Biosystems
cDNA Library Arrayed by: The J.M.A.G.B. Consortium (LINTL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be

Page 2

FEATURES

ORIGIN

505 GGTATATTACAAATTACATTGCATACTGATATATCCACGCCCTTTGAGGATGTAAC 564

Db 11 GCTAATTTCAGGTTCTGCAGTGCAACCAGCGCTTATCCGCTGCCGCTAGAGCACGTCAAC 70

Db 71 CTCAGTCTGATCACCGAGTTCGAGAGGAGTTCCGACATTCGCATCGGATACTCCGGA 130

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	52
--	---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	----

[illegible][illegible][illegible]

100

RESULT 2

DEFINITION **DAYS / JO** **HOMO sapiens** **INITIUM** **HOMO sapiens** **CENH** **CENH** **CCDC68**

5-PRIME; mRNA sequence.

RECORDS	ESI:
SOURCE	Homo sapiens (human)

REFERENCE 1 (pages 1 to 1200)
Mammalia; Eutheria; Primates; Carchiniini; Homiidae; Homo.

JOURNAL
Unpublished (2001)

50

ORIGIN

	Query March	Score 79.4	DB 13	Length 1200
Beet Local Similarity	30.2%	Pred. No. 0.00013		
Matches 259	Conservative 167	Mismatches 423	Indels 10	Gaps 1

86 AACCGGCAAGTTCCAACTTGAAGAATTAACAATTTCAG 145

2025

466 TTTTWWAKAAAAARAAIWIWWCAAAAAANWIIWWAAARGRAKGAIIIIIIIAHWWG 525

Db 526 GGAGRTAHTWTITTTTTTTTAAHWAATWATAATTTTWWTWAWAAAAASACGAK 585

Db 586 ATWTTTCKKTTATWKKRAGATWTTT TTTTWTAARGDWKWTTWTTT TTTT 645

Db 646 TTTTITGTTTWWRDGARTAKATAKWTGTTTWTAAWAGGAARKBRATWTWTAAAWA 705

706 WWAGAAARCGAGAGRAWTTTTTTATKAGARCATWTTAWAWMTATRAAGARGADTTT 765

Do 766 TTATATWTTATTTTWWAAGATDKAAAAAWWWTTTWTAAAAAA--TT 815

Db 816 WWWAGARAAKATTTWAAWAAAAAAWRAKWAATATATATTTATATATAAKARAAA 875

D5 876 ATAAATTAARARRWKGAATAAATAATTWATAAAAAAATTWAWWR 935

Db 936 A A A R A W W A A A A G A R A G A W A A W A T T W T T T T T A A T A T R W W A A A A A W T T T T T W T 995

1.

Db 996 WAAAAAAMNDTTTATATATATATATATATAGAAAATATADATATATATATATADAAAAGAA 1055
 QY 746 ATGACACTGAAACATCTTTGTATATGGGGTCAGCTGTTCGAAAAATCTTAGGTCATAA 805
 Db 1056 WATKDDAGAGATAAAWTTTTTTTTTTTTTTTAAAAAAGAAAAATTTTATAAAAAATAAR 1115
 QY 806 GTAAAGTGTATACAGCTCAGAAAAGAAATATAATCTGACAGAAAGTCTATTATAG 865
 Db 1116 AAAAAATTTTTAAAAAAAAAAAAAAAAAAGAKKDAAWAGAAATTAATAAABAAAAAAGWZAR 1175
 QY 866 CTAAACAGAGATAAAAA 884
 Db 1176 AAAAAAATAAAAAATATARA 1194

	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
RESULT 3	BC050955	BC050955	1435 bp	mRNA	linear	HTC 10-APR-2003	
LOCUS	BC050955	Danio rerio, clone IMAGE:5915349, mRNA.					
DEFINITION	BC050955	Danio rerio, clone IMAGE:5915349, mRNA.					
ACCESSION	BC050955						
VERSION	BC050955.1	GI:28747823					
KEYWORDS	HTC						
SOURCE	Danio rerio (zebrafish)						
ORGANISM	Danio rerio						

REFERENCE
AUTHORS
TITLE
JOURNAL
Submitted (10-APR-2003)
Direct Submission
Submitted (10-APR-2003)
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA

REMARK	COMMENT
NIH-MGC Project URL: http://mgc.nci.nih.gov	Contact: MGC help desk

Tissue Procurement: Dr. Sumio Sugano
cDNA Library Preparation: Dr. Sumio Sugano
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNI)
DNA Sequencing by: Genome Sequence Centre,
BC Cancer Agency, Vancouver, BC, Canada
info@bcsc.bc.ca
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
Susuma Chao, Readman Chiu, Chris Flell, Elin Gardlan, Ran Guin,
Leticia Hsiao, Martin Krywinski, Reta Kutsche, Oliver Lee, Soo
Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Jacquelin
Sess, Pawan Pandoh, Anna-Lisa Prabhu, Parvaneh Seedi, Jacqueline
Schein, Duane Smalls, Michael Smith, Lorraine Spence, Jeff Stott,
Michael Thorne, Miranda Tsai, Nataja van den Bosch, Jill Vardy,
George Yang, Scott Zuydamyuan, Marco Marra.

Clone distribution: MCC clone distribution info can be found through the I.M.A.G.E Consortium/ILNI at: <http://image.llnl.gov>
Series: IRAP Plate: 14 Row: 1 Column: 23
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, similarity but not identity to protein
This clone has the following problem: no cloning site / microclonon.

FEATURES
source

```

/organism="Danio rerio"
/mol_type="mRNA"
/db_xref="taxon:7955"
/clone="IMAGE:5915344"
/tissue="embryo body, adult male"
/clone_lib="DSHGano SUD adult male"
/lab_host="DH10B"
/note="Vector: pMT185-FL3"

```

ORIGIN

Query Match 7.48; Score 78.8; DB 11; Length 1435;

Best Local Similarity 48.4%; Pred. No. 0.00013;
Matches 218; Conservative 0; Mismatches 232; Indels 0; Gaps 0;

0Y	479	TTTTTAAATAATAAAAGTTCCGGTGGTAAATATTACAAATTTCATTGCAATTCGAAT	538
Db	609	TTTTTAAAGATTGTAAAGAGACAAATAGAATCTTTGCATCTCCGAGGTGCAAGTGCAT	668
0Y	539	ATCCAGAGCCCTTTGAGGATATGAACCTTAATGCTATTAAATGATTTAAAAAACACTTCC	598
Db	669	ATCCACTTGAACCTGAGAGACGTCAACTCGGGGTTATCAAGAAATATCAGAAAGATTCC	728
0Y	599	CTAAGAAATACATAGCTTCTCTGATCATCTTCAGCGGTTTATGCAAGCTATTTCCGGCGG	658
Db	729	CAGACATTCCCATGGTACTCGGATCGGGATAGAAATGGAGATCMAATCAACAGTCCGGGCGAG	788
0Y	659	TGCCTTATGGAATAACTTTTAATTAATAAAACATTTCACTTAGATTAATCTATGTCTGGCC	718
Db	789	TTGCTCTGGGAGCCAAAGTTGTGAGAGCGGCATGTGACCTCGAATTAAGAGCTGGAAAGCCA	848
0Y	719	CAGATCATTTGGCTCATAGACACCTGATGAATCTGTAACATCTTTGTAATTTGGGTCAGST	778
Db	849	GTAATCATATGACAGCTCTCTCTGGACACAGAAGAGCTGGCTGAGCTGGTGCATCCGCA	908
0Y	779	GTAATGAAAAATCTTTAGGTTCAAAATAGTAAGTGGTAAACAGTTCAGAAAGCAAGATA	838
Db	909	TTGTGGAGAGAGCTTTTGGGCACTGGAATGAAAGCGATGCTGCCCTGTGAGGTTGCCATGCC	968
0Y	839	AAATCGTAGCAAGAGCTTATTTATAGCTTAAACAGAGATTAATAAAAAAGGTAGGTTT	898
Db	969	ATGATTAAGCTCGGAGAGTGGTGGTGGCAAGACCTTCATTTCAAAAAGGTACAGAACTCA	1028
0Y	899	CAGAAAAAATATATACAACAAGAAAGCCTG	928
Db	1029	CCCTGACATGCTGACATGTGAATTAAGCTG	1058

RESULT 4	CD327087	876 bp	mRNA	linear	EST 28-MAY-2003
LOCUS	CD327087				
DEFINITION	CD327087				
ACCESSION	MGNCOURT_14147023	NICHD_XGC_Eyel	Xenopus laevis	CDNA clone	
VERSION	IMAGE:6948035.5	mRNA sequence.			
KEYWORDS	CD327087				
SOURCE	CD327087.1	GI:31091418			
ORGANISM	Xenopus laevis	(African clawed frog)			
	Xenopus laevis				

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1 (bases 1 to 876)
NIH-MGC <http://mgc.ncl.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Daniela S. Gerhard, Ph.D.

National Cancer Institute / NIH
Bldg. 31 Rm10A07 Bethesda, MD 20892
Email: cga@bbs-remail.nih.gov
Tissue Procurement:
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium {LIML}
DNA Sequencing by: Agencourt Bioscience Corporation
clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LIML at:
<http://image.llnl.gov>
plate: LIML14561 row: k column: 10
High quality sequence stop: 685.

FEATURES
SOURCE

```
/organism="Xenopus laevis"  
/mol_type="mRNA"  
/db_xref="taxon:8355"  
/clone="IMAGE:6948035"
```

```
/clone="IMAGE:6948035"
```

ORIGIN

RESULT 5	CA550431	LOCUS	DEFINITION
	CA550431	508 bp	mRNA
	C03030806-5N	NIA Mouse BlascoGyb	cDNA Library (Long)
	CNA clone NIA:C03030E06	IMAGE:30027893	5', mRNA sequence.
			EST 19-NOV-2002

ACCESSION	CAS500431	GI:25094538
VERSION	CAS500431.1	
KEYWORDS	EST.	
SOURCE	Mus musculus (house mouse)	
ORGANISM	Mus musculus	

REFERENCE	1 (bases 1 to 508)
AUTHORS	Piao, Y., Kargul, G.J., Dudenkula, D. B., Qian, Y., Tanaka, T., Luo, A. and
TITLE	Ko, M.S.H.
JOURNAL	Systematic Analyses of NIA Mouse Blasticyst cDNA Library (Long)
COMMENT	Unpublished (2001)
	Other ESTs: C0830E06-3

Department of Geriatrics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6620, USA
Email: cdna@gsun.gic.nia.nih.gov
Plate: C0830 row: E column: 06
Seq primer: M13 Reverse
High quality sequence stop: 508

```

Location/Qualifiers
1..508
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57Bl/6d"
/db_xref="tae8t:C0830B06-5N"
/db_xref="taxon:10090"
/cdon="NTA:C0830B06 IMAGE:30027893"
/tissue_type="BlasccCyst"

```

```

/def stage="1.5-opc"
/lab host="Dhl01a"
/cclone lib="NIA Mouse Blastocyst cDNA Library (long)"
/ncore=vector: pSPORT1 (Invitrogen); site_1: SalI; Site_2:
NotI; Mouse cDNA project by the Laboratory of Genetics,
National Institute on Aging (NIA), Intramural Research
Program, NIH (http://jgsun-grc.nia.nih.gov/cDNA/). This is
a long-transcript enriched cDNA library (Ref. Genome Res.
11: 1553-1558 (2001)). [PMID: 11541991]. Total RNAs were
extracted from a pool of 20 blastocysts. Double-stranded
cDNAs were synthesised with an Oligo(dT) primer

```

5'-GAGCTAGTTCAGATCCGACGCGCGCCCTCTTTT-3') from total RNA, treated with T4 DNA polymerase, and purified by ethanol-precipitation. The cDNAs were ligated to *l*-one-linker *l*-Sal-I, purified by phenol/chloroform, and separated from free linkers by centrifugation. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Taq polymerase (Takara) with a primer Sal-4-S. The products were purified by phenol/chloroform and centrifugation. The cDNAs were digested with SalI and NotI enzymes and cloned into SalI/NotI site of pSPORT1 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by the standard chemical method. The average insert size is about 2.2 kb. The library was constructed by Yulan Zhao (NIA)."

```

ORIGIN
Query Match      6.8%      Score 71.8; DB 14; length 508;
Best Local Similarity 50.9%      Pred. No. 0.0034;
Matches 200; Conservative 0; Mismatches 187; Indels 6; Gaps 1

```

QY	12	ATATATCGTTCCTGAAATGATGTGACCAACCATATGATGTGATATTTGCAGAGGAAT	71
Db	40	ATTATATATTCGAGAAATGAGTGGCAATCATATATCATCTATAGATATAGCATTTAAAT	99
QY	72	GATATTTAAACCCAAAGGCGCGGTGTTAATGCGGTAAATTTCCAAATTTAAAGTCGA	131
Db	100	TATTGAACTCTCTGCAGAAATGTGGGCCCAAGCAATTTAACTACAAACATATACGCCGA	159
QY	132	TAAA-----TTAATTTGCGATTTGCACTTAAGGACAGATATCAATTTTTAAACACAG	185
Db	160	TACATACACCTTAATATCATAGAGGTGACCTTTTGATATTCAGATGAAAGCTTTTGG	219
QY	186	AGAAATTAGAAATCTCAGTTAGAAATGACAAAAAGCTTGAATGAAGTATGACATTATCT	245
Db	220	GAAAAACAGAAATTTGATGAACTATATCAAGAAACATATCTCATGGAAATGGCATGA	279
QY	246	CCATCTATAGGAATATGCAAGTCAGTTTAAATTAGATGTTTTCTTACCCCTTTTGACGA	305
Db	280	AGCTGTTATTTAAATGTCGACCTGATPAACATATCATTTATTTTTCTACACCTTTTGATGA	339
QY	306	AGACTCTATGATTTTTTTAGCATCTTTGAAACAAAAATATGSAAAATCCCTTCAGGTGA	365
Db	340	CAGTGGCGGTGATTTTTTTAGAAAGTTAAATTGGCCAGCCTATPAATAAATGCGGCTTTTGA	399
QY	366	GTTATTTGAATTTACCGTATCTTGAAAAAATATAGC	398
Db	400	AAATACTATTTTACAGTTATTTAAAAAGGTAC	432

RESULT 6
BI377301

Thu May 13 11:53:12 2004

us-09-930-440b-7.rst

Page 5

LOCUS B1377301 545 bp mRNA linear EST 26-AUG-2003
DEFINITION BFLG3 001104 Amphioxus 5-6 hrs cDNA library (Name convention: BFLG
or MEMGp98) Branchiostoma floridae cDNA clone MEMGp98L1399 5',
mRNA sequence.
ACCESSION B1377301 GI:30912237
VERSION B1377301.1 GI:30912237
KEYWORDS EST.
SOURCE Branchiostoma floridae (Florida lancelet)
ORGANISM Branchiostoma floridae
Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
Branchiostoma.
REFERENCE 1 (bases 1 to 545)
Panopoulou,G., Hennig,S., Groth,D., Krause,A., Poustka,A.J.,
Herwig,R., Vingron,M. and Lehrach,H.
New evidence for genome-wide duplications at the origin of
vertebrates using an amphioxus gene set and completed animal
genomes
Genome Res. 13 (6A), 1056-1066 (2003)
JOURNAL MEDLINE
PUBMED 12793946
COMMENT Contact: Panopoulou G
Laboratory 145, dept.Lehrach
Max-Planck-Institut fuer Molekulare Genetik
Inster.63-73, D-14195 Berlin, Germany
Tel: +49 30 8413 1235
Fax: +49 30 8413 1128
Email: panopoulou@gen.mpg.de
The library was characterized by oligonucleotide fingerprinting
(ONFP) to reduce sequencing redundancy. According to the ONFP
procedure, clones giving the same hybridization pattern with a
battery of 200 8mer oligonucleotides are grouped into clusters. One
clone per cluster is selected for sequencing. The size of each
cluster is an indicator of the frequency of a transcript in the
analysed library. The cluster size as well the coordinates of the
rest of the clones assigned to the same fingerprint cluster as the
clone from which the above EST is generated is available at the
amphioxus project site at: <http://www.molgen.mpg.de/amphioxus/>
Database and filters are distributed via the Resource Center/Primary
PCR primers
FORWARD: 5' CCCGAGCTTACCTTAACTTCCGCG 3' (M13RSP)
BACKWARD: 5' GCTATACGCGACCTGCGGAGAGGGAGTGTG 3' (M13FSF)
Insert Length: 1200 Std Error: 0.00
Seq primer: 5'-CCGCTCCGGAATCCCGGT-3' pSPORT3/86
High quality sequence stop: 545.
Location/Qualifiers
1..545
/organism="Branchiostoma floridae"
/mol_type="mRNA"
/db_xref="taxon:7739"
/clone="MEMGp98L1399"
/tissue_type="whole embryo"
/dev_stage="5-6 hrs (gastrula stage)"
/lab_host="R. coli, XL1 blue"
/clone_lib="Amphioxus 5-6 hrs cDNA library (Name
convention: BFLG or MEMGp98)"
/note="Vector: pSPORT1; Site 1: SalI; XbaI, EcoRI (5');
Site 2: NotI, BamHI, HindIII (3'); OligodT primed and
directionally cloned in pSPORT vector using a NotI
(5'-GACTAGTTCTAGATCGGAGCGCGGCC (T)15-3' and a SalI 5'-
TCGACCCACCGCTCCG-3' adapters (Gibco BRL)."

Db 159 GCGGCTGTGGGACCCGTACAGAGAAATATGCGGGCGAGAAAGCCCTGTGTGTCATGCA 218
Qy 525 TTGCAATAGTGAATATCCAAAGCCCTTGAAGATGTAACCTTAATGCTATTAATATT 584
Db 219 GTGTACACACATACACCCCTCCCTCCACCGACGTAATCGAGGTATTAATCTTTT 278
Qy 585 GAAAAACACTTCCCTAAGATATACATAGAGCTTCTGTATCATTTAGCGGGTTTATGC 644
Db 279 CAGAGAAAGTTTCCCAAGCCATCATGTGTATCTGTGTGACAGATCGATATCCAT 338
Qy 645 AGCTATGTGGCGGCTGCTTATGAAATACCTTTATGAAAAATTGACCTTTAGATTA 704
Db 339 CTCCTGGAGCGTTCAGTCTGGGTCCAAAGTCGTGGAGCGTCACTGACAA 398
Qy 705 ATCTATGTGGCGGCTGCTTATGAAATACCTTTATGAAAAATTGACCTTTAGATTA 764
Db 399 GTCTGGAAGGCTGTGACCGCGGCTGCTGACGTTGATGAGCTTAAGAGATCTTG 458
Qy 765 TATTGGGTCAGGTGTGTAATAATCTTTAGGTTCAATATGTAAG 811
Db 459 TGCACATCCGATTTGTGAGAGCGCTCTGGGCAAACTATCAAG 505
RESULT 7
LOCUS B1185296/c
DEFINITION UNL-P-FN-cu-a-06-0-UNL.51 UNL-P-FN Sus scrofa cDNA clone
ACCESSION B1185296
VERSION B1185296.1 GI:14659705
KEYWORDS EST.
SOURCE Sus scrofa (pig)
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE 1 (bases 1 to 715)
Caetano,A.R., Johnson,R.K. and Pomp,D.
Generation and sequence characterization of a normalized cDNA
library from swine ovarian follicles
Mamm. Genome 14 (1), 65-70 (2003)
JOURNAL MEDLINE
PUBMED 12532269
COMMENT Contact: Pomp, D
Department of Animal Science
University of Nebraska, Lincoln
Lincoln, NE 68583-0908, USA
Tel: 402 472 6416
Fax: 402 472 6362
Email: dpomp@unl.edu
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to verify it as a clone from the
normalized porcine ovarian follicles library
Seq primer: M13 -29
POLY(A)-yes.
Location/Qualifiers
1..715
/organism="Sus scrofa"
/mol_type="mRNA"
/strain="University of Nebraska, Lincoln Swine Selection
lines"
/db_xref="taxon:9823"
/clone="UNL-P-FN-cu-a-06-0-UNL"
/dev_stage="ADULT"
/lab_host="DH10B (Life Technologies)"
/clone_lib="UNL-P-FN"
/note="Vector: pTZ19-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; The UNL-P-FN
library is a normalized library representing porcine
ovarian follicles, ranging between 2.0 to 10.0 mm in
diameter, collected during 7 days of the follicular phase

of the pig estrous cycle. This library was derived from the library UNL-P-2. The tag is a string of 5-6 nucleotides present between the Not I site and the oligo-dT track. The library was constructed as described by Ronald, Lennon and Soares, Genome Research 6: 791-806, 1996.

TAG_TISSUE=porcine ovarian follicles
TAG_LIB=UNL-P-FN
TAG_SEQ=CACACT

ORIGIN

Query Match 6.5%; Score 68.6; DB 12; Length 715;
Best Local Similarity 48.7%; Pred. No. 0.011;
Matches 185; Conservative 0; Mismatches 195; Indels 0; Gaps 0;

521 TACATTCATATGATGATATCCACGCCCTTGGAGATGAAACCTTAATGCTATTATG 580
572 TCCAGTGCACCGCGGTACCCACTCCACCTGAGACGTCACCTGCGTCACTCGG 513
581 ATTGAAAAACCTTCCTTACGATTAACATAGGCTTCTGATCTCTAGCGGCTTT 640
512 AATATCAGAAAGCTTCTTCCGATTCATCCATAGGATATTCTGGCATGAAACGCGATG 453
641 ATGACCTATATGCGCGGTGCTTATGATTAATCTTTATGAAAACATTCATCTTATG 700
452 CAAATCTGTGCTGCGGTGCTGCTGCGACCAAGCTCTGAGCTCACTAATCTTGG 393
701 ATAAATCTATGCTGCGCCAGATCATTTGGCTCAATGAACTGAACTGAAACATC 760
392 ACAAGCTTGGAGGAGGAGTACCTCGCCCTCGTGAAGCTGAGAGCTGCGCGAGC 333
761 TTGTATGCGGCTGAGGTGTTGAAATATCTTAGATCAAAAGTAAAGTGTATGACG 820
332 TGTGCGGTCTGCGCCCTGCTGAGAAAGCCCTGAGCTCCCAACCAACGACCTGCTG 273
821 CTTCAGAAAGAGAAATAAATGCTAGCAAGAAAGTCTATATAGCTAAACAGAGATTA 880
272 CCTGTGAGATGCGCTCAACGAGAAAGCTGGCAAGTGTGCTGCGCAAGTGAATTC 213
881 AAAAGTGAAGTTTTTCA 900
212 CGAAGGCACTGTTCTTAAC 193

RESULT 8 868 bp mRNA linear EST 27-MAY-2003
CD301354
LOCUS AGENCOURT 14234689 NICHG_XGC_Eyel Xenopus laevis cDNA clone
DEFINITION IMAGE:6959626 5', mRNA sequence.
ACCESSION CD301354
VERSION CD301354.1 GI:31081149
KEYWORDS EST.
SOURCE Xenopus laevis (African clawed frog)
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
Xenopodidae; Xenopus.
1 (bases 1 to 868)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Daniela S. Gerhard, Ph.D.
Office of Cancer Genomics
National Cancer Institute / NIH
Bldg. 31 Rm10A07 Bethesda, MD 20892
Email: cgabs-remail.nih.gov
Tissue Procurement:
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov

Plate: LLM14591 row: n column: 09
High quality sequence start: 16
High quality sequence stop: 706.
Location/Qualifiers
1..868
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="IMAGE:6959626"
/dev stage="adult"
/lab host="DH10B (phage-resistant)"
/note="Organ: eye; Vector: pCMV-SPORE6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2.3 kb. Constructed by Life
Technologies. Note: This is a Xenopus Gene Collection
(XGC) library."

FEATURES

source

ORIGIN

Query Match 6.5%; Score 68.6; DB 14; Length 868;
Best Local Similarity 45.9%; Pred. No. 0.01;
Matches 365; Conservative 0; Mismatches 409; Indels 21; Gaps 3;

16 ATCGTGTGAATGTTGCAACCATATGATGTTGATATTGCAAGAAATGATA 75
72 ATCATGCGGAGATCGGCGAGAACCAAGGCGCATGAAATGCCAAGAAATGATC 131
76 TTAAGGCCAAAGAGCCGGTGTATGACGATTAATTCCAAACATTTAAAGCTGATTA 135
132 CTAAAGCGAAGGATGGGTGCGATGTCGAAAGTTGAGAAAGGAACTGAAATAC 191
136 TTAATTCAGCTATTCGACCTAAGCAGATGATCAATTAATAAACACAGAGAAATTAGA 195
192 AAGTTCATTAAGAAAGCTCTGAGAGGACATTAATCTCCACACTGCTGGG---GAAA 248
196 TCTCATGAAATGACAAAAGAGCTGAAATGAAGTGAACGATTATCTCATATG 255
249 ACATATGAGAGGACCAACGCGCAGCTGAGATGACCCATGATACCGGAGATACG 308
256 GAATATGAGGACGATTTAAATTTAGATGTTTCTTCTACCCCTTTCAGAGACACTAT 315
309 AAGTATGAAAGAGAGATTTATTTTACAGACTGTGATGAGAGATGCGCATG 368
316 GATTTTTCAGATCTTTGAAACAAAATAATGAAATCCCTTCAAGTATGAT 375
369 GAATTCCTTCATGAGCTGGCGTCCCTTTTCAAGGGGTGAGGGAGACCAATATAC 428
376 TTACGCTATCTTGAATAAATAGCAGAGCTTCCGATCCCTGATTAAGAAATATATCA 435
429 TTCAATATTTAAAGACCTGCAAAAAGAGTGGCCCT-----ATGATATATCA 479
436 ACAGAGATGCTACTATGATGATGATTAAGATCAAGCTGTGTTCTATTTTAAATATTA 495
480 AGCGGCGATGAGCTGCTGGAACCAATGCGTGTGTTATGATCTGTGAACAAGTAA- 538
486 GTTCCGGTGGTGAATTTATTAATTTATGATGCAATCTGAAATTCGCAAGCCCTTGA 555
539 -----TCCAACTTTTGTCTTACAGGCAACATGATGATCCACTTACCAAGG 590
556 GATGTAACCTTAATGCTATTAATGATTTGAAAAACACTTCCCTTAAGATTAACATAGC 615
591 GAGCTCAACCTGTGATGATTAAGAGATTTCACTTCCAGATATCCCAATTTGGC 650
616 TTCTGTATCATTTAGCGGGTTTATGACGATATGCGGGGTGCTTATGATTAATCT 675
651 TACTGTGTATGAAATCTGCAATGCTATCTCAATGAGGAGCTGTGCTATGAGGACCA 710
676 TTATGAAAAACATTTCACTTATGATTAATGATGCTGAGCCGATCATTTGGCTCA 735
711 GTGTGAGAGCGCATGTAACCTGACAAATCTGGAAGGAATGATCCAAAGCTTGG 770
736 ATGAGACTGATGAACCTGAACATTTTGTATGAGGCTGAGGTGTTGAAAAATCTTTA 795

Thu May 13 11:53:12 2004

us-09-930-440b-7.rst

Page 7

Db 771 CTGAGCCGACGATGATGACAGATATGATGACGCTCCATCATGACTGTGAGAAAAAGCAATG 830
QY 796 GGTCAAAATAGTAAA 810
Db 831 GGGCTCACTGTCAAA 845

RESULT 9:
BM095216/c 676 bp mRNA linear EST 20-NOV-2001
E3J1811.X1 Sugano SJD adult male Danio rerio cDNA clone
IMAGE:5410364 3 similar to TR:Q9VG74 Q9VG74 CG5232 PROTEIN. ;
mRNA sequence.

ACCESSION BM095216 GI:17024182
VERSION EST
KEYWORDS Danio rerio (zebrafish)
SOURCE
ORGANISM

REFERENCE
AUTHORS Clark, M., Johnson, S.L., Lehrach, H., Lee, R., Li, F., Marra, M.,
Boddy, S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,
Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y.,
Persson, B., Swaller, T., Gibbons, M., Page, D., Harvey, N., Schurk, R.,
Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R.,
Waterston, R. and Wilson, R.
Washu Zebrafish EST Project 1998
Unpublished (1998)

TITLE
JOURNAL
COMMENT
Contact: Stephen L. Johnson
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810

Email: zbrafish@watson.wustl.edu
Library constructed by Dr. Sumio Sugano and Dr. Koichi Kawakami.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution information can be found through the I.M.A.G.E.
Consortium/BLMT, send email to: info@image.llnl.gov
High quality sequence stop: 525.

FEATURES
SOURCE
Location/Qualifiers
1..676

/organism="Danio rerio"
/mol_type="mRNA"
/db_xref="taxon:7955"
/clone="IMAGE:5410364"
/sex="male"
/tissue_type="whole body"
/dev_stage="adult"
/lab_host="DH10B (phage resistant)"
/clone_lib="Sugano SJD adult male"
/note="Vector: pME18-Flu3; Site 1: DraIII (CACCATGTG);
Site 2: DraIII (CACTGTGTG); 1st strand cDNA was primed
with an oligo(dT) primer (ATGCGCCCTTTTCTTTTCTTTT);
double-stranded cDNA was ligated to a DraIII adaptor
(CTGGGCGCTGTG), digested and cloned into distinct DraIII
sites of the pME18-Flu3 vector (5' site CACGTGTG, 3' site
CACCATGTG). XhoI should be used to isolate the cDNA
insert. Size selection was performed to exclude fragments
<1.5kb. Library constructed and donated by Dr. Sumio
Sugano (University of Tokyo Institute of Medical Science).
Custom primers for sequencing: 5' end primer
CTTCGCTCTTAAGCTGCG and 3' end primer
GCACCTGACGCTCAGACACA."

ORIGIN

Query Match 6.4%; Score 68.2; DB 12; Length 676;
Best Local Similarity 55.2%; Pred. No. 0.013;
Matches 133; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

Db 675 TCGAGTGCACAAAGTCATATCCACTGGAACCTGAGAGCGTCAACCTGCGGTATCCAGG 616
QY 581 ATTGGAAGAAACCTCCCTTAAGATATACATAGCGCTCTGTATCATTTAGCGGGTTT 640
Db 615 AATATGAGAAAGATTCACAGCATTCCTCATCGGTACTGGAAGTGGGATCA 556
QY 641 ATGACCTATTGCGGGGCGGCTTATGATATCTTTATGAAAAACATTTACCTTAG 700
Db 555 ACATCACTGCGGGGAGTTGCTTGGGAGCCAAAGTTGAGCGGCACTGACCCCTGG 496
QY 701 ATTAATCTATGCTGCGCCAGATCATTTGCGCTCATATGAACTGTATGAAATC 760
Db 495 ATTAGAGCTGGAAGGACATGATATGACAGCGCTCTGGAACCAAGAGCTGAGAG 436
QY 761 T 761
Db 435 T 435

RESULT 10
CF360452/c 673 bp mRNA linear EST 25-AUG-2003
LOCUS CF360452
DEFINITION 821784 MARC 3P1G Sus scrofa cDNA 3', mRNA sequence.
ACCESSION CF360452
VERSION CF360452.1 GI:34158171
KEYWORDS
SOURCE
ORGANISM

Sus scrofa (pig)
Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
Smith, T.P.L., Preking, B.A., Ford, J.J., Vallet, J.L., Fox, J.,
Wise, T.A., Noneman, D.J., Wray, J.B. and Keeler, J.W.
A second set of porcine ESTs from a pooled-tissue normalized
library
Unpublished (2003)

TITLE
JOURNAL
COMMENT
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390

Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called with phred v0.020425.c and
trimmed with the aid of the trim_alt option. Vector identified with
cross match v0.990329.
plate: SRG8013 row: K column: 12
Seq primer: TAGAGGACACAGTCGAGG.
Location/Qualifiers
1..673

/organism="Sus scrofa"
/mol_type="mRNA"
/db_xref="taxon:9823"
/tissue_type="pooled"
/lab_host="DH10B"
/clone_lib="MARC 3P1G"
/note="Vector: pCDNA3.1; Site 1: EcoRI; Site 2: NotI;
Library made with RNA pooled from multiple tissues
including brain, liver, muscle, placenta/endometrium,
ovary, testes, and bone marrow."

ORIGIN

Query Match 6.4%; Score 67.6; DB 14; Length 673;
Best Local Similarity 46.6%; Pred. No. 0.017;
Matches 251; Conservative 0; Mismatches 285; Indels 3; Gaps 1;

QY 521 TACATTCGATATGATATATCCAGCGCCCTTGAGAGATGTAACTTAATCTATTATG 580
Db 560 TCCAGTGCACCAAGCGGTATCCATCCATGAGATGAGAGTCAACATGCGTATCTCCG 501
QY 581 ATTGGAAGAAACCTCCCTTAAGATATACATAGCTTCTGTATCATTTAGCGGGTTT 640
Db 500 AATATGAGAAAGATCTTTCATGACATTCCTCCATAGGATTTCTGGGCATGAAGCGGATAG 441

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OW nucleic - nucleic search, using sw model

Run on: May 11, 2004, 07:15:29 ; Search time 81.495 Seconds

(without alignments)
7211.402 Million cell updates/sec

Title: US-09-930-440B-7

Perfect score: 1059
Sequence: 1 atgagtaataatataatcgt.....atcgagaaacaataatcgt 1059

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 682709 seqs, 27747546 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database:

Issued Patents NA: *
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq: *
2: /cgn2_6/ptodata/2/ina/5A_COMB.seq: *
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq: *
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq: *
5: /cgn2_6/ptodata/2/ina/PTCUTS_COMB.seq: *
6: /cgn2_6/ptodata/2/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	166.2	15.7	1664976	4 US-08-916-421B-1	Sequence 1, Appl1
2	105.2	9.9	11474	4 US-09-495-406-1	Sequence 1, Appl1
3	63	5.9	1080	4 US-09-516-143A-3	Sequence 3, Appl1
4	63	5.9	1230	4 US-09-620-312D-1035	Sequence 1035, Ap
5	56.8	5.4	7218	4 US-08-232-463-14	Sequence 14, Appl
6	52.6	5.0	5152	4 US-10-204-708-73	Sequence 73, Appl
7	48.4	4.6	640681	4 US-09-720-988-1	Sequence 1, Appl1
8	47.8	4.5	640681	4 US-09-720-988-1	Sequence 1, Appl1
9	47.2	4.5	11049	4 US-10-204-708-21	Sequence 21, Appl
10	45.6	4.3	15016	4 US-09-601-198-60	Sequence 60, Appl
11	44.6	4.2	5222	4 US-08-936-165A-12	Sequence 12, Appl
12	44.4	4.2	6866	4 US-10-204-708-20	Sequence 20, Appl
13	44.4	4.2	26000	4 US-09-843-376-10	Sequence 10, Appl
14	43.8	4.1	2049	4 US-09-107-532A-845	Sequence 845, App
15	43.2	4.1	658	4 US-08-998-416-595	Sequence 595, App
16	43.2	4.1	19124	4 US-08-487-826B-13	Sequence 13, Appl
17	42.2	4.0	984	4 US-09-107-532A-2527	Sequence 2527, Ap
18	42	4.0	6243	4 US-09-056-075-1	Sequence 1, Appl1
19	42	4.0	580073	4 US-08-545-528D-1	Sequence 1, Appl1
20	41.8	3.9	356	4 US-08-520-678A-22	Sequence 22, Appl
21	41.8	3.9	1194	4 US-08-897-126-22	Sequence 22, Appl
22	41.8	3.9	9646	4 US-08-811-586-1	Sequence 1, Appl1
23	41.8	3.9	9646	4 US-08-811-586-1	Sequence 1, Appl1
24	41.8	3.9	9646	4 US-09-034-756-1	Sequence 1, Appl1
25	41.2	3.9	960	4 US-08-916-421B-1	Sequence 1, Appl1
26	41.2	3.9	960	4 US-09-134-001C-2716	Sequence 2716, Ap
27	41.2	3.9	2126	3 US-08-235-836C-75	Sequence 75, Appl

ALIGNMENTS

28	41	3.9	1635	4 US-09-601-198-10	Sequence 10, Appl
29	41	3.9	269223	4 US-09-596-002-41	Sequence 41, Appl
30	40.8	3.9	1686	4 US-09-134-001C-1354	Sequence 1354, Ap
31	40.6	3.8	832	4 US-09-621-976-2813	Sequence 2813, Ap
32	40.6	3.8	2463	4 US-09-134-000C-3032	Sequence 3032, Ap
33	40.6	3.8	580073	4 US-08-545-528D-1	Sequence 1, Appl1
34	40.4	3.8	13086	4 US-08-956-172B-16	Sequence 16, Appl
35	40.2	3.8	252	4 US-09-540-236-1345	Sequence 1345, Ap
36	40.2	3.8	5152	4 US-10-204-708-74	Sequence 74, Appl
37	40	3.8	1353	4 US-09-601-198-37	Sequence 37, Appl
38	40	3.8	2448	4 US-08-526-964-2	Sequence 2, Appl1
39	40	3.8	2448	4 US-08-546-617-2	Sequence 2, Appl1
40	40	3.8	2448	4 US-09-031-897-2	Sequence 2, Appl1
41	40	3.8	3203	4 US-09-257-770-1	Sequence 1, Appl1
42	40	3.8	99500	4 US-09-798-096-10	Sequence 10, Appl
43	39.8	3.8	2355	3 US-09-371-913A-6	Sequence 6, Appl1
44	39.8	3.8	2355	4 US-09-967-805-6	Sequence 6, Appl1
45	39.8	3.8	6317	4 US-10-204-708-11	Sequence 11, Appl

```

RESULT 1
US-08-916-421B-1
Sequence 1, Application US/08916421B
Patent No. 6503729
GENERAL INFORMATION:
APPLICANT: Built et al.
TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanoco
Patent No. 6503729
TITLE OF INVENTION: jannaschii
FILE REFERENCE: PB275
CURRENT APPLICATION NUMBER: US/08/916,421B
CURRENT FILING DATE: 1997-08-22
PRIOR APPLICATION NUMBER: US 60/024,428
PRIOR FILING DATE: 1996-08-22
NUMBER OF SEQ ID NOS: 3
SOFTWARE: PatentIn version 3.1
SEQ ID NO 1
LENGTH: 1664976
TYPE: DNA
ORGANISM: Methanococcus jannaschii
FEATURE:
NAME/KEY: misc feature
LOCATION: (28222)..(28222)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (28257)..(28258)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (84773)..(84773)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (84808)..(84808)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (84812)..(84812)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98120)..(98120)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98159)..(98159)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98239)..(98239)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98266)..(98266)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98343)..(98343)
OTHER INFORMATION: n equals a, t, c, or g

```

```

OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (103998)..(103998)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (148948)..(148948)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (163385)..(163385)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (191989)..(191989)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (191995)..(191995)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (231980)..(231980)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (234187)..(234187)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (234220)..(234220)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (234814)..(234814)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (309398)..(309398)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (309418)..(309418)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (312837)..(312837)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (312993)..(312993)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (319226)..(319226)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (559167)..(559167)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (559241)..(559241)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (600992)..(600992)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (622708)..(622708)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (657081)..(657081)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (657203)..(657203)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (674435)..(674435)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (682442)..(682442)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (713652)..(713652)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (741684)..(741684)
OTHER INFORMATION: n equals a, t, c, or g

```

```

NAME/KEY: misc_feature
LOCATION: (779455)..(779455)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (779676)..(779676)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (855539)..(855539)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (871519)..(871519)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1084830)..(1084830)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1096846)..(1096846)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1119881)..(1119881)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1130881)..(1130881)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1310988)..(1310988)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1313224)..(1313224)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1349473)..(1349473)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1349491)..(1349491)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1470091)..(1470091)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1569020)..(1569020)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1602912)..(1602912)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1603734)..(1603734)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1637998)..(1637998)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1664854)..(1664854)
OTHER INFORMATION: n equals a, t, c, or g
US-08-916-421B-1

Query Match      15.7%; Score 166.2; DB 4; Length 1664976;
Best local Similarity 52.8%; Pred. No. 1.5e-28;
Matches 438; Conservative 0; Mismatches 373; Indels 18; Gaps 3;

Cy 180 CACGAGATTTCAGTCTGATGAGAAAGCAAAACCTGGAATGAGTACGCA 239
Db 465620 CATAGCAAAAATCAGAAATATATGACTATTTAAAGTTTGAATCTGCAAGAGCA 465679
Cy 240 TTATCTCATCTATGAGATATGAGATGAGTCACTTTAAATTGAGATGTTTCTACCCCTTT 299
Db 465680 ATTCTATGAGCTAAATGAAATGCAAAATGGAATATATGTTATCTCAACACATT 465739
Cy 300 TGACGAGACTCTATGATTTTTCATCTTTGACATCTTTGAAACAAAATATGGAATCCCTC 359
Db 465740 AGATTTAATATGATATATATTAATTAATGAATGTCCTGCATTTAAATTCCTC 465799
Cy 360 AGGTGATTTATGAAATTTACCGTATCTTGAAAAAATGCAAGCTCCGATCCCTGATAA 419

```

Db 465800 TCGTGAATTAACCTTTATCCCTTATTTAGAAAAAGTCGCAAAA-----CAGGCA 465850
Qy 420 GAAATATATCATATCAAGAAATGGCTACTATGATAGATATTAAGATAGTCTGTTTAT 479
Db 465851 GCGGTGATTTTATTTACAGAAATCTGATTTGGGAAATTTGGGAGACATTAAGT 465910
Qy 480 TTTTATTAATAATAAAGTCCGTTGGTATATTAACAATTTACATTCGAATACTGAATA 539
Db 465911 TTTAATAATATATGATGACAGG-----ATATATTTATTTAGCTTTGATTTCAATCTTA 465964
Qy 540 TCCAGAGCCCTTGGATGATTAACCTTAATGCTTTATATGATTTGAATAAAACCTTCC 599
Db 465965 CCCAACCCCTTATATGAGATGCAATTAAGCTATTAACCTTGAATAAGTATTTTCA 466024
Qy 600 TAAGATTAACATAGCTTCTCTGATCATCTTACGCGGTTTATGACGCTATTCGCGCGT 659
Db 466025 TATCCCT---GTGGATATTTCTGACCATATGGAATATCGCCCAAGTATTTCTGT 466081
Qy 660 GCGTTATGCAATTAATTTTATTTGAAAACTTTCCTTATGATTAATCTATGTCGCCC 719
Db 466082 TCCCTTAGAGCGGATGTTATTTAGAGACACTTACCTTAGATTAATAATATGGAAGTCC 466141
Qy 720 AGATCATTTGGCCTCAATAGAACCTGATGAACCTGTAATCTTTGATTTGGGCTCAGGTG 779
Db 466142 TGTATATGCTTTGTGACGACACCCGAGAAATTTAGAAATGTTATATACATTAAGTT 466201
Qy 780 TGTGAAAAATCTTTAGGTTCAATATATATATATGATTAAGTGTTCACCTTCAAGAAAGATATTA 839
Db 466202 AGTTGAAAAAATGCTTGAAGTGGGAAAGCATACCAATGCTTCTGAAAGACGTTAT 466261
Qy 840 AATGCTACCAAGAAAGCTATATATGCTTAAACAGATTAATAAAAGGTGAGCTTTTTC 899
Db 466262 TGTATAGCAAGAAAGATATTTGTCAGAAAGATTTAAAGAAAGGAAATATCTTAG 466321
Qy 900 AGAAAAAATATTAACAACAAAAAGACTGTATATGATATCACTCCGATGAGTGTATTA 959
Db 466322 TGTATATATATTTTATTTAAAGACCGGAGAGATTTGAAACAAAGATTTTAGCAT 466381
Qy 960 TTTATTTGGTAAATTTGACAGACACACTTATTTCCATGATTAATAA 1008
Db 466382 AATATTAACAGAAAAATCAAAAAAGCATTAAGAGATGATATATTA 466430

RESULT 2

US-09-495-406-1
Sequence 1, Application US/09495406
Patent No. 6503744
GENERAL INFORMATION:
APPLICANT: Gilbert, Michel
APPLICANT: Wakarchuk, Warren W.
APPLICANT: National Research Council of Canada
TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
FIDE REFERENCE: 019633-030110US
CURRENT FILING DATE: 2000-01-31
PRIOR APPLICATION NUMBER: US 60/118,213
NUMBER OF SEQ ID NOS: 35
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 11474
TYPE: DNA
ORGANISM: Campylobacter jejuni
FEATURE:
OTHER INFORMATION: 11.5 kb PCR product from C. jejuni OH4384
OTHER INFORMATION: including LOS biosynthesis locus
US-09-495-406-1
Query Match 9.9%; Score 105.2; DB 4; Length 11474;
Best Local Similarity 49.1%; Pred. No. 4.3e-15;
Matches 479; Conservative 0; Mismatches 458; Indels 39; Gaps 6;

Qy 17 TCGTTCGAAATTTGTTGCAACCATATATGATATGTTGATATTCAGAAAGATATAT 76
Db 6979 TCGGCTCGTAAATATAGCATTAATCATATATGCGAGTTAGAACTGATTAATATGTTAG 7038
Qy 77 TAAAGCCAAAGAGCGCGGTATTAATGACATTAATTTCCAAAGTTTAAAGCTATAAT 136
Db 7039 ATGCAAGCTTTAGACAGGTGCTTAAATTTAAAGATCAAAACCAATCTTGAAGATG 7098
Qy 137 TAAATTCAGCTATGACCTTAAGCAGATCAATTAATAAAACAGAGGAATTAAGT 196
Db 7099 AGATGAG-----TAAAGCCCTTAAAGAAATTTCTGTAATGCAAAATAA 7146
Qy 197 CTCAGTTAGAAATAGCAAAAAAGCTTAATTAAGTATGAGTATCTCATCTAATG 256
Db 7147 GCATTTATGATATATGCAAAATATGCTTATGATTTAAAGATGACCTTACCTTAAG 7206
Qy 257 AATATGACATGATTAATTTATGATGTTTTCACCCCTTTGACGAAGACTATATG 316
Db 7207 AATACAGAAATATTTAGTCTTGTATCTTAGCAACCTTTTCTGACAGTGA 7286
Qy 317 ATTTTATGATCTTTGAAACAAAAATATGAAATCCCTTCAGGTGATTAATG 376
Db 7267 ACCGCTTGAAGATATGAGATGATGCTTTTAAAGATTGTTGACGTGAGTATATAT 7326
Qy 377 TACGTTCTTGAATAAATATGCAAGCTTCCGATCCCTGATTAAGAAATATCATCA 436
Db 7327 ATCCGCTTATTAACCATAG-----CAGCTTTAAAGCCTTATGATAGTGA 7377
Qy 437 CAGGAATGCTACTATATATGATATTAAGACAGTCTGTTCTATTTATTAATAATTAAG 496
Db 7378 CAGGATGATATGATTAATTAAGATTAATTAACCACTGTAATAATCTTTAGACATATAA 7437
Qy 497 TCCGCTTGTATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 556
Db 7438 TTCCCTTGT-----TTAATGCAACACCAATCTTTACCAACCCCGATATC 7488
Qy 557 ATGTAACCTTAATGCTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 616
Db 7489 TTTGATGATTAAGCTATGCTTAATTAATTAATTAATTAATTAATTAATTAATTA 7545
Qy 617 TCTGTGATCTTACGCGGTTTATGACGATATGCGGGGCTTATGATTAATCTT 676
Db 7546 TACGACACCAACAGATATCTGCTGTTTAAAGTGGGTGACCTGCTGCTGCTG 7605
Qy 677 TTTTGAATAAATTTTCACTTATATATATATATATATATATATATATATATATATAT 736
Db 7606 TCGTTGAAGACATTTTATGATATATATATATATATATATATATATATATATATAT 7665
Qy 737 TAGAAGCTGATGAATGAACATTTTGTATGAGGCTCAGGTGTTGAAAAATCTTTAG 796
Db 7666 TCGATACACAGGCTTAAAGAGCTTATTAACAAAGTGAACAAATGCTATATATGAG 7725
Qy 797 GTTCAAT 856
Db 7726 GAATTAATTAAGTAAATTAAGACCTTAAAGCAAGCAAGTCAATGATTTGCTTGG 7785
Qy 857 CTATTAATGCTTAATAC---AGATTAATAAAGGTGAGTTTTCAGAAAAATATTA 913
Db 7786 CAAGCTATGATGATTAAGATATTAATAAAGCGAAGTTTATCTATGATATATATTT 7845
Qy 914 CAACAAAAAGACCTG---TATGCTATGCTGATGAGTGTATATATATATATATAT 970
Db 7846 GCGTTAAAGACCTGAGCTTGTGATATATATATATATATATATATATATATATAT 7905
Qy 971 AATATGCAAGCAAG 986
Db 7906 AAAAGCATTAAGAG 7921

RESULT 3

US-09-516-143A-3
Sequence 3, Application US/09516143A

Thu May 13 11:53:11 2004

us-09-930-440b-7.rnt

Page 4

Patent No. 6333182
GENERAL INFORMATION:
APPLICANT: Human Genome Sciences, Inc.
TITLE OF INVENTION: Human Glycosylation Enzymes
FILE REFERENCE: P505PCT
CURRENT APPLICATION NUMBER: US/09/516,143A
PRIORITY FILING DATE: 2000-03-01
PRIORITY APPLICATION NUMBER: 60/122,409
PRIORITY FILING DATE: 1999-03-02
NUMBER OF SEQ ID NOS: 6
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 3
LENGTH: 1080
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(1077)
US-09-516-143A-3

Query Match 5.9%; Score 63; DB 4; Length 1080;

Best Local Similarity 47.8%; Pred. No. 1.2e-05; Mismatches 200; Indels 0; Gaps 0;

Matches 183; Conservative 0; Mismatches 200; Indels 0; Gaps 0;

518 TATTACATTGCAATACGATATATCCAGCCCTTGAAGATGTAACCTTAATGCTATTA 577
542 TCTTCAGATGTAACGAGCATACCGCTCCAGCTGAGAGACCTCACTCGGGTCACT 601
578 ATGATTTGAAAAAACAATCTCCCTAGATTAACATAGGCTCTCTGATATCTTACGGCT 637
602 CGGATATGCAAGCTCTTCTTCTGATCCATCCATAGGGTATCTCGGGATGTAACAGGCA 661
638 TTATGACGATATGCGCGGCTGCTTATGGAATTAATTTATGAAAAACATTTCACTT 697
662 TAGGATATCTGTGGCCGAGTGGCTCTGGGGGCCAAGGTGTGGAAGTCAATTAATCT 721
698 TAGTAAATCTATGCTGCGCCAGATCACTTTGGGCTCAATGAACTGATTAATCTGAAAC 757
722 TGACAGACCTGGAAGGGAGTGAACCACTCGGCTCGCTGAGCTGGAAGAACTGGCCG 781
758 ATCTTTGATTTGGGCTCAGTGTGTGTAATAATCTTTAGTTCAATATGTAAGTGTGA 817
782 AGCTGTGCGGTCACTGCTCTGTGAGCGGTGCTCCCAACCAAGAGCTGC 841
818 CAGCTTCAGAAAGGAATTAATCTGAGCAAGAACTGTTATTAAGCTTAACAGAGA 877
842 TGCCCTGTGAGATGGCTGCAATGAGAACTGGGCAAGTCTGTGGTCCCAAGTGAAGA 901
878 TAAAAAAGGTGAGGTTTTC 900
902 TTCGGAGGCACTTCTAACA 924

RESULT 4

US-09-620-312D-1035

Sequence 1035, Application us/09620312D

Patent No. 6569652

GENERAL INFORMATION:

APPLICANT: Tang, Y. Tom
APPLICANT: Liu, Chenghua
APPLICANT: Asundi, Vinod
APPLICANT: Zhang, Jie
APPLICANT: Ren, Feiyun
APPLICANT: Chen, Rui-hong
APPLICANT: Zhao, Qing A.
APPLICANT: Weinman, Tom
APPLICANT: Xue, Aidong J.
APPLICANT: Yang, Yonghong
APPLICANT: Wang, Jian-Rui
APPLICANT: Zhou, Ping
APPLICANT: Ma, Yundong
APPLICANT: Wang, Dunrui
APPLICANT: Wang, Zhiwei

APPLICANT: John Tillinghast
APPLICANT: Dymnac, Radoje T.
TITLE OF INVENTION: No. 6569652el Nucleic Acids and
FILE REFERENCE: Polypeptides
FILE REFERENCE: 784CIP2B
CURRENT APPLICATION NUMBER: US/09/620,312D
PRIORITY FILING DATE: 2000-07-19
PRIORITY APPLICATION NUMBER: 09/552,317
PRIORITY FILING DATE: 2000-04-25
PRIORITY APPLICATION NUMBER: 09/488,725
PRIORITY FILING DATE: 2000-01-21
NUMBER OF SEQ ID NOS: 1105
SOFTWARE: pt. fl. genes Version 1.0
SEQ ID NO 1035
LENGTH: 1230
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (119)..(1198)
US-09-620-312D-1035

Query Match 5.9%; Score 63; DB 4; Length 1230;

Best Local Similarity 47.8%; Pred. No. 1.2e-05; Mismatches 200; Indels 0; Gaps 0;

Matches 183; Conservative 0; Mismatches 200; Indels 0; Gaps 0;

518 TATTACATTGCAATACGATATATCCAGCCCTTGAAGATGTAACCTTAATGCTATTA 577
660 TCTTCAGATGTAACGAGCATACCGCTCCAGCTGAGAGACCTCACTCGGGTCACT 719
578 ATGATTTGAAAAAACAATCTCCCTAGATTAACATAGGCTCTCTGATATCTTACGGCT 637
720 CGGATATGCAAGCTCTTCTTCTGATCCATCCATAGGGTATCTCGGGATGTAACAGGCA 779
638 TTATGACGATATGCGCGGCTGCTTATGGAATTAATTTATGAAAAACATTTCACTT 697
780 TAGGATATCTGTGGCCGAGTGGCTCTGGGGGCCAAGGTGTGGAAGTCAATTAATCT 839
698 TAGTAAATCTATGCTGCGCCAGATCACTTTGGGCTCAATGAACTGATTAATCTGAAAC 757
840 TGACAGACCTGGAAGGGAGTGAACCACTCGGCTCGCTGAGCTGGAAGAACTGGCCG 899
758 ATCTTTGATTTGGGCTCAGTGTGTGTAATAATCTTTAGTTCAATATGTAAGTGTGA 817
900 AGCTGTGCGGTCACTGCTCTGTGAGCGGTGCTCCCAACCAAGAGCTGC 959
818 CAGCTTCAGAAAGGAATTAATCTGAGCAAGAACTGTTATTAAGCTTAACAGAGA 877
960 TGCCCTGTGAGATGGCTGCAATGAGAACTGGGCAAGTCTGTGGTCCCAAGTGAAGA 1019
878 TAAAAAAGGTGAGGTTTTC 900
1020 TTCGGAGGCACTTCTAACA 1042

RESULT 5

US-08-232-463-14/C

Sequence 14, Application US/08232463

Patent No. 5670367

GENERAL INFORMATION:

APPLICANT: DORNER, F.
APPLICANT: SCHREIFLINER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:

Page 5

```

P. PRIOR APPLICATION NUMBER: PCT/EP01/03971.
P. PRIOR FILING DATE: 2001-04-06
P. PRIOR APPLICATION NUMBER: DE 10019058.8
P. PRIOR FILING DATE: 2000-04-06
P. PRIOR APPLICATION NUMBER: DE 10019173.8
P. PRIOR FILING DATE: 2000-04-07
P. PRIOR APPLICATION NUMBER: DE 10032529.7
P. PRIOR FILING DATE: 2000-06-30
P. PRIOR APPLICATION NUMBER: DE 10043826.1
P. PRIOR FILING DATE: 2000-09-01
P. NUMBER OF SEQ ID NOS: 98
P. SEQ ID NO 73
P. LENGTH: 5152
P. TYPE: DNA
P. ORGANISM: Artificial Sequence
P. FEATURE:
P. OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-73

Query Matchh          5.0%; Score 52.6; DB 4; Length 5152;
Best Local Similarity 47.4%; Pred. No. 0.0042;
Matches 251; Conservative 0; Mismatches 274; Indels 4; Gaps 3.

QY 119 CATTAAAGCTGATAAATTAAATTTCAGCTTATGCACCTTAAGGCGAGCTATCAAAATAAAA 178
      |||||
DB 1116 CATTAAATATATCTACTTCACTTCATTTCCATTAACCTTACTTCCCTTTAAATTATTAATAATAAA 1057

QY 179 ACACAGAGAAATTAGATCTCAGTTAGAAAATGACAAAAGAGCTGGAATGAAGATGACG 238
      |||||
DB 1056 AACTATATATATAAATG-CATCTATTTTAAAGCTTAAAAAAGCAAAATCTCTATATATATA 998
      |||||

```

```

QY      239  ATTATCCGATCTAATGGAATGCACTGAGTTAAATTTAGATGTTTTCACCCCTT  298
Db      997  AAAAAATATTTAAATAATTTTAATTAACCTATTAAATTAATTTATCTTTATACCTTT  938
QY      299  TTGACGAAAGACTATTTGATTTTTTGACATCTTTGAAACAAAAATATGAAAAATCCCTT  358
Db      937  ATATTATTAATCTCTTAATATTTTCATATTTAAAAATTTAAAAACAAC-TTTTAATTI  879
QY      359  CAGGTGATATGAAATTTACCGATCTGAAAAAATAGCCACACTCCGATCCCTGATA  418
Db      878  CAGAAACTTAATACCTTTAAAAAATTAATATATTAATTTATTAACATATATTTA  819
QY      419  AGAAATTAATCATATCAACAGGATGGCTACTATTTGATGAGATGAAAAAGCTGTGTTCTA  478
Db      818  TAATAATAATATATAAACAATTACTACTATTTATATATTAATAAAACTTACTCAT  759
QY      479  TTTTATTAATTAATAAGTCCGGTGTGA--TATTAACATATTAATGATGCATATCTGA  536
Db      758  ATCTTTCAAACTTTCAATTCACCTCACGACCACTACTAAATTCCTTATTCAAAAATTA  699
QY      537  ATATCCACGCCCTTGAGGATGTAACCTTAATGCTATTATGATTTGAAAAAACAACCT  596
Db      698  TAATTTAACCTAATAATTTTTTAAATAAAAAATTAATAATATCTTTTAAATATATT  639
QY      597  CCTTAAGATTAACATAGCCTTCTCTGATCATCTTAGCGGCTTTATAGCA  645
Db      638  ATACAAAAATTAACAACATATTTTATTTATTAATTAATTTTTCATATCA  590

RESULT 7
US-09-790-988-1
: Sequence 1, Application US/09790988
: Patent No. 6632935
: GENERAL INFORMATION:
: APPLICANT: SHIGENOBU, SHUJI
: APPLICANT: WATANABE, HIDEKI
: APPLICANT: WATANABE, MASAHIRA
: APPLICANT: HATTORI, YOSHIYUKI
: APPLICANT: SAKAKI, YOSHIYUKI
: TITLE OF INVENTION: GENE/DNA OF BACTERIAL SYMBIONT OF APHIDS
: FILE REFERENCE: 081356/0159
: CURRENT APPLICATION NUMBER: US/09/790,988
: CURRENT FILING DATE: 2001-02-23

```

Thu May 13 11:53:11 2004

us-09-930-440b-7.rn1

Page 6

PRIOR APPLICATION NUMBER: JP2000-107160
PRIOR FILING DATE: 2000-04-07
NUMBER OF SEQ ID NOS: 7
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 640681
TYPE: DNA
ORGANISM: Buchnera sp.
US-09-790-988-1

Query Match
Best Local Similarity 4.6%; Score 48.4; DB 4; Length 640681;
Matches 189; Conservative 0; Mismatches 201; Indels 0; Gaps 0;

QY 266 TCAGTTAAATTTAGATGTTTCTACCCCTTTGACGAAGCTCATTTGATTTTAC 325
DB 50171 TATTCAAAATATGATTTTCTTAAATATCTATGATTAATAAATGATTTTAAA 90230
QY 326 CATCTTGAAACAAATATGGAATCCCTCAGCTGATTAATGATTTACCATC 385
DB 90231 GATTAATGATTAATAAATATGGAATATGATTAATAAATATTTTGGATTT 90290
QY 386 TTGAAAAATAGCCAGCTCCGATCCGATTAAGAAATATATCATCAAGAAATG 445
DB 90291 TAACTATATAGATTTTCAACAAAAAAGATTAAGATTTATCTTTATTAAT 90350
QY 446 CATCATTTGATGATTAATAAGCTGCTTTCTATTTTAAATTAATTAAGTCCGATG 505
DB 90351 TTCCCTTTCTTTTAAATTTTCTTACCAATTTTAAAGATTAAGTAAATTTTC 90410
QY 506 GATTAATATCAATATTTACATGCAATGATTAATCAACGCCCTTGAGATTAACC 565
DB 90411 AAAGTAAATATTAATTTATTTGAACTTAATCAACCTTAAGTTTGGCATGATCT 90470
QY 566 TTAATGCTTAATGATTTGAAAAAGCTTCCCTTAATTAATTAAGTCTTCTGATC 625
DB 90471 TTAATAATATTTTATTTGAAATCAACATTAATTAAGTAAATTAATTTTCTCT 90530
QY 626 ATCTAGCGG 635
DB 90531 ATTTTATAG 90540

RESULT 8
US-09-790-988-1/c
Sequence 1, Application US/09790988

PATENT NO. 6632935
GENERAL INFORMATION:
APPLICANT: SHIGENOBU, SHUJI
APPLICANT: MATSUMOTO, HIDEKI
APPLICANT: MATSUMOTO, HIDEKI
APPLICANT: MATSUMOTO, HIDEKI
TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
FILE REFERENCE: 081356/0159
CURRENT APPLICATION NUMBER: US/09/790,988
PRIOR FILING DATE: 2001-02-23
PRIOR APPLICATION NUMBER: JP2000-107160
NUMBER OF SEQ ID NOS: 7
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 640681
TYPE: DNA
ORGANISM: Buchnera sp.
US-09-790-988-1

Query Match
Best Local Similarity 4.5%; Score 47.8; DB 4; Length 640681;
Matches 118; Conservative 0; Mismatches 117; Indels 0; Gaps 0;

QY 111 ATTCGAATTTAAAGCTGATTAATTAATTTGATTAATGACCTTAAGGAGATTA 170
DB 353778 ATGCAAAAATGATTTCTTTTAAATCTTCTATTAATTTCAAGAAAAAGATTTT 353719

QY 171 AATTAATAACAGAGAGATTTAGATCTAGTAAATGACAAAAAGCTTGAATGAA 230
DB 353718 AATTAATAACAGAGATTTAGATCTAGTAAATGACAAAAAGCTTGAATGAA 353659
QY 231 GATGAGATTAATCTCATCTTAATGAAATGAGTCAAGTTAAATTTAGATTTTTC 290
DB 353658 TGTATTTTCTTTTGTATTAATTTTAAATTAATTAATTAATTTTGAATTT 353599
QY 291 TACCCCTTTGACGAGACTGATTAATTTTAAATTTTAAATTTTAAATTTTAA 345
DB 353598 AATTAATTAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 353544

RESULT 9
US-10-204-708-21/c
Sequence 21, Application US/10204708
PATENT NO. 6677731
GENERAL INFORMATION:
APPLICANT: OLEK, Alexander
APPLICANT: FIEBENROCK, Christian
APPLICANT: FIEBENROCK, Christian
TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
FILE REFERENCE: 5013.1012
CURRENT APPLICATION NUMBER: US/10/204,708
PRIOR FILING DATE: 2003-05-06
PRIOR APPLICATION NUMBER: PCT/EP01/03971
PRIOR FILING DATE: 2001-04-06
PRIOR APPLICATION NUMBER: DE 10019058.8
PRIOR FILING DATE: 2000-04-06
PRIOR APPLICATION NUMBER: DE 10019173.8
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: DE 10032529.7
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: DE 10043826.1
PRIOR FILING DATE: 2000-09-01
NUMBER OF SEQ ID NOS: 98
SEQ ID NO 21
LENGTH: 11049
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-21

Query Match
Best Local Similarity 4.5%; Score 47.2; DB 4; Length 11049;
Matches 178; Conservative 0; Mismatches 218; Indels 0; Gaps 0;

QY 96 TGTATATGATTAATTTCAAACTTAATTAAGCTGATTAATTTTAAATTTTAAATTTTAA 155
DB 8049 TGTATATGATTAATTTCAAACTTAATTAAGCTGATTAATTTTAAATTTTAAATTTTAA 7990
QY 156 TAAAGCAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 215
DB 7989 TACTTAAATTAATTTCAAACTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 7930
QY 216 TAAAGCTGATTAATTTCAAACTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 275
DB 7923 TACTTAAATTAATTTCAAACTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 7810
QY 276 TTTAGATTTTCTTACCCCTTTGACGAGACTGATTAATTTTAAATTTTAAATTTTAAATTTTAA 335
DB 7869 TTTAGATTTTCTTACCCCTTTGACGAGACTGATTAATTTTAAATTTTAAATTTTAAATTTTAA 7810
QY 336 AAAAAAATTAATTAATTTCAAACTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 395
DB 7809 AAAAAAATTAATTAATTTCAAACTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 7750
QY 396 AAAAAAATTAATTAATTTCAAACTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 455
DB 7749 AAAAAAATTAATTAATTTCAAACTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 7690

Thu May 13 11:53:11 2004

us-09-930-440b-7.rn1

Page 8

RESULT 12

US-10-204-708-20/c
Sequence 20, Application US/10204708
Patient No. 6677731
GENERAL INFORMATION:
APPLICANT: OLEX, Alexander
APPLICANT: PIEPERROCK, Christian
APPLICANT: BERLIN, Kurt
TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
TITLE OF INVENTION: by Assessing DNA Methylation
FILE REFERENCE: 5013.1012
CURRENT APPLICATION NUMBER: US/10/204.708
CURRENT FILING DATE: 2003-05-06
PRIOR APPLICATION NUMBER: PCT/EP01/03971
PRIOR FILING DATE: 2001-04-06
PRIOR APPLICATION NUMBER: DE 10019058.8
PRIOR FILING DATE: 2000-04-06
PRIOR APPLICATION NUMBER: DE 10019173.8
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: DE 10032529.7
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: DE 10043826.1
PRIOR FILING DATE: 2000-09-01
NUMBER OF SEQ ID NOS: 98
SEQ ID NO 20
LENGTH: 6866
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-20

Query Match 4.2%; Score 44.4; DB 4; Length 6866;
Best Local Similarity 48.8%; Pred. No. 0.35;
Matches 120; Conservative 0; Mismatches 126; Indels 0; Gaps 0;

QY 108 AAAATTCGAACATTTAAAGCTGATAATTAATTTGAGTATTCAGCTTACAGGACAGAGTA 167
DB 948 AAAATTCCTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 889
QY 168 TCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 227
DB 888 TAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 829
QY 228 GAGATGATGAGATTAATTCATCTCATCTATGATGATGATGATGATGATGATGATGAT 287
DB 828 AAAACATTCACCAAACTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 769
QY 288 TTCTACCCCTTTGACGAGATCTATGATTTTGAATCTTTGAAACAAAATATG 347
DB 768 CGCTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 709
QY 348 GAAAT 353
DB 708 TAAAT 703

RESULT 13

US-09-843-376-10/c
Sequence 10, Application US/09843376
Patient No. 656132
GENERAL INFORMATION:
APPLICANT: C. Frank Bennett
APPLICANT: Andrew T. Walt
TITLE OF INVENTION: ANTISENSE MODULATION OF INTERFERON GAMMA RECEPTOR 1 EXPRESSION
FILE REFERENCE: PUS-0234
CURRENT APPLICATION NUMBER: US/09/843.376
CURRENT FILING DATE: 2001-04-26
NUMBER OF SEQ ID NOS: 88
SEQ ID NO 10
LENGTH: 26000

TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
US-09-843-376-10

Query Match 4.2%; Score 44.4; DB 4; Length 26000;
Best Local Similarity 44.3%; Pred. No. 0.48;
Matches 180; Conservative 0; Mismatches 226; Indels 0; Gaps 0;

QY 116 AAACATTTAAAGCTGATAATTAATTTAGCATTTGACCTTACAGGACAGATGCAATTA 175
DB 2415 ATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 24056
QY 176 AAACACGAGAGATTTAGATCTCATGTAATGAAATGACAAAAGCTTGAAATGAAATG 235
DB 24055 ATATTAATTAATTTATTAATTTATTAATTTATTAATTTATTAATTTATTAATTT 23996
QY 236 AGATTAATCTCATGTAATGAAATGAGTCACTTAATTTAGATGTTTCTTACCC 295
DB 23995 TATATAATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 23936
QY 296 CTTTGAAGAGAGCTCTATGATTTTGAATCTTGAATCTTGAATCTTGAATCTTGAAT 355
DB 23935 AACTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 23876
QY 356 CTTCAAGTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 415
DB 23875 TATATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 23816
QY 416 ATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 475
DB 23815 TATATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 23756
QY 476 CTAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 521
DB 23755 ATATTAATTCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 23710

RESULT 14

US-09-107-532A-845
Sequence 845, Application US/09107532A
Patient No. 6583275
GENERAL INFORMATION:
APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 7310
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107.532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Axinello, Pamela Denise
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781) 893-5007

TELEFAX: (781) 893-8277
 INFORMATION FOR SEQ ID NO: 845:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2049 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: circular
 MOLECULE TYPE: DNA (genomic)
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 ORIGINAL SOURCE:
 ORGANISM: Enterococcus faecium
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: (B) LOCATION 1..2049
 SEQUENCE DESCRIPTION: SEQ ID NO: 845:
 US-09-107-532A-845

Query Match 4.1%; Score 43.8; DB 4; Length 2049;
 Best Local Similarity 50.2%; Pred. No. 0.35; Indels 0; Gaps 0;
 Matches 108; Conservative 0; Mismatches 107;
 QY 48 TAGTGTGATATGCAAGAGAAATGATTTAAAGCCAGCGTGTATGAGT 107
 DB 1329 TCGGATTCGATTTAAAAAATCGATTTCAAGAGAAATGATGTTGTTATTTCTAC 1388
 QY 108 AAAATTCGAACATTTAAAGCTAATTAATTTGCTAGTCTAGCTAAGCAAGTA 167
 DB 1389 ATCTATTTTCCAGAGATACGGTGAAATATATTTAGTCTCCCGTACTGCTAGAGA 1448
 QY 168 TCAATTAATAAACAAGAGATTTAGATCTCAGTTGAATGACAAAAAGCTTGAAT 227
 DB 1449 TGAAGTCAGACAGCTTAAGAAGCTTTTGCATATTTGAGAGAAAAAGCAAAAAAT 1508
 QY 228 GAAGTATGACGATTTATCTCCATCTAATGGAATATG 262
 DB 1509 TACGTATTAACAGCAATTAACGATGATGATAGAG 1543

RESULT 15
 US-08-998-416-595/C
 Sequence 595, Application US/08998416
 Patent No. 6239264
 GENERAL INFORMATION:
 APPLICANT: Philippsen, Peter
 APPLICANT: Pohlmann, Rainer
 APPLICANT: Steiner, Sabine
 APPLICANT: Mohr, Christine
 APPLICANT: Wendland, Jürgen
 APPLICANT: Knechtle, Philipp
 APPLICANT: Reibschung, Corinne
 TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPTII
 TITLE OF INVENTION: AND USES THEREOF
 NUMBER OF SEQUENCES: 1152
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: No. 6239264artis Corporation
 STREET: 3054 Cornwallis Road
 CITY: Research Triangle Park
 STATE: No. 6239264th Carolina
 COUNTRY: USA
 ZIP: 27709
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/998,416
 FILING DATE: 24-DEC-1997
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: CH 0016/97
 FILING DATE: 31-DEC-1996

ATTORNEY/AGENT INFORMATION:
 NAME: Meigs, J. Timothy
 REGISTRATION NUMBER: 38,241
 REFERENCE/DOCKET NUMBER: PF/5-30306/R/CGC1976
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 919-541-8587
 TELEFAX: 919-541-8689
 INFORMATION FOR SEQ ID NO: 595:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 658 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 ORIGINAL SOURCE:
 ORGANISM: PAG1408BP
 US-08-998-416-595

Query Match 4.1%; Score 43.2; DB 3; Length 658;
 Best Local Similarity 43.0%; Pred. No. 0.36; Indels 0; Gaps 0;
 Matches 210; Conservative 0; Mismatches 278;
 QY 105 AGTAAATTCGAACATTTAAAGCTGATTAATTAATTTGACGCTATGCACTTAAGCAG 164
 DB 652 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 593
 QY 165 GTATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 224
 DB 592 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 533
 QY 225 AATGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 284
 DB 532 GAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 473
 QY 285 TTTTCTACCCCTTTTGAAGAGCTCTATGATTTTGAAGCTTTTGAAGCAAAAAAT 344
 DB 472 ATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 413
 QY 345 ATGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 404
 DB 412 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 353
 QY 405 TCCGATCCCTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 464
 DB 352 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 293
 QY 465 ACAGTCTGTTTATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 524
 DB 292 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 233
 QY 525 TTGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 584
 DB 232 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 173
 QY 585 GAAAAAC 592
 DB 172 AAAGCATC 165

Search completed: May 11, 2004, 07:23:09
 Job time: 94.495 secs

Thu May 13 11:53:12 2004

us-09-930-440b-7.rnpb

Page 1

GenCore Version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OW nucleic - nucleic search, using sw model

Run on: May 11, 2004, 11:46:35 / Search time 490,708 Seconds
(without alignments)
9776.195 Million cell updates/sec

Title: US-09-930-440B-7
Sequence: 1 atgagcattatataatcgt.....aatgagacaaataattatcg 1059

Scoring table: IDENTITY NUC
Gapop 10'-0, Gapext 1.0

Searched: 2941586 segs, 2264955651 residues

Total number of hits satisfying chosen parameters: 5883172

Minimum DB seg length: 0

Maximum DB seg length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA:*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1059	100.0	1059	US-09-930-440B-7	Sequence 7, Appl1
2	326.4	30.8	17276	US-09-870-759-83	Sequence 83, Appl1
3	326.4	30.8	17276	US-09-751-708A-83	Sequence 83, Appl1
4	324.8	30.7	25020	US-10-192-280-1	Sequence 1, Appl1
5	105.2	9.9	11474	US-09-816-028A-1	Sequence 1, Appl1
6	105.2	9.9	11474	US-10-303-161-1	Sequence 1, Appl1
7	105.2	9.9	11474	US-10-303-118-1	Sequence 1, Appl1
8	105.2	9.9	11474	US-10-303-128-1	Sequence 1, Appl1
9	105.2	9.9	11474	US-10-303-134-1	Sequence 1, Appl1
10	105.2	9.9	11474	US-10-303-162-1	Sequence 1, Appl1
11	70.4	6.6	11223	US-09-882-227-489	Sequence 489, Appl1
12	63	5.9	1080	US-09-984-205-3	Sequence 3, Appl1
13	63	5.9	1080	US-09-930-440B-5	Sequence 5, Appl1
14	63	5.9	1230	US-10-037-270-1035	Sequence 1035, Ap

15	63	5.9	1230	US-10-117-722-1035	Sequence 1035, Ap
16	63	5.9	1268	US-10-264-237-1192	Sequence 1192, Ap
17	57.2	5.4	410	US-10-106-698-2649	Sequence 2649, Ap
18	56	5.3	1023	US-10-335-977-2864	Sequence 2864, Ap
19	54.6	5.2	3673778	US-10-312-841-1	Sequence 1, Appl1
20	52.8	5.0	6394	US-10-821-613-239	Sequence 239, Appl1
21	52.6	5.0	5152	US-10-204-708-73	Sequence 73, Appl1
22	52.6	5.0	5152	US-10-240-589C-115	Sequence 115, Appl1
23	52.2	4.9	942	US-10-335-977-2863	Sequence 2863, Ap
24	50.4	4.8	624	US-10-027-632-232815	Sequence 232815, Ap
25	50.4	4.8	624	US-10-027-632-232815	Sequence 232815, Ap
26	50.4	4.8	624	US-10-027-632-232815	Sequence 232815, Ap
27	50.4	4.8	624	US-10-027-632-232815	Sequence 232815, Ap
28	49.4	4.7	6061	US-10-311-455-114	Sequence 114, Appl1
29	48.4	4.6	640681	US-09-790-988-1	Sequence 1, Appl1
30	48	4.5	1326	US-10-027-632-123225	Sequence 123225, Ap
31	48	4.5	1326	US-10-027-632-123225	Sequence 123225, Ap
32	48	4.5	1326	US-10-027-632-123225	Sequence 123225, Ap
33	48	4.5	1326	US-10-027-632-123225	Sequence 123225, Ap
34	48	4.5	11650	US-10-221-714A-481	Sequence 481, Appl1
35	47.8	4.5	572	US-10-027-632-194960	Sequence 194960, Ap
36	47.8	4.5	572	US-10-027-632-194960	Sequence 194960, Ap
37	47.8	4.5	640681	US-09-790-988-1	Sequence 1, Appl1
38	47.2	4.5	11049	US-10-204-708-21	Sequence 21, Appl1
39	47.2	4.5	11049	US-10-311-455-641	Sequence 641, Appl1
40	47.2	4.5	11049	US-10-240-589C-27	Sequence 27, Appl1
41	47	4.4	21354	US-10-221-714A-512	Sequence 512, Appl1
42	46.6	4.4	1136	US-09-938-842A-3011	Sequence 3011, Ap
43	46.6	4.4	1136	US-09-938-842A-3011	Sequence 3011, Ap
44	46.6	4.4	9504	US-10-240-453-281	Sequence 281, Appl1
45	45.8	4.3	3673778	US-10-312-841-2	Sequence 2, Appl1

ALIGNMENTS

RESULT 1

US-09-930-440B-7

Sequence 7, Application US/09930440B

Patent No. US20020142386A1

GENERAL INFORMATION:

APPLICANT: Becandough et al.

TITLE OF INVENTION: Engineering Intracellular Sialylation Pathways

FILE REFERENCE: P5509P2

CURRENT APPLICATION NUMBER: US/09/930,440B

CURRENT FILING DATE: 2001-08-16

PRIOR APPLICATION NUMBER: 60/227,579

PRIOR FILING DATE: 2000-08-25

PRIOR APPLICATION NUMBER: 09/516,793

PRIOR FILING DATE: 2000-03-01

PRIOR APPLICATION NUMBER: 60/169,624

PRIOR FILING DATE: 1999-12-08

PRIOR APPLICATION NUMBER: 60/122,582

PRIOR FILING DATE: 1999-03-02

NUMBER OF SEQ ID NOS: 8

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 7

LENGTH: 1059

TYPE: DNA

ORGANISM: Escherichia coli

FEATURE:

NAME/KEY: CDS

LOCATION: (1) .. (1041)

US-09-930-440B-7

Query Match 100.0%; Score 1059; DB 9; Length 1059;

Best Local Similarity 100.0%; Pred. No. 5.2e-214;

Matches 1059; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 ATGAGTAAATATATATATATCGTGTCTGAATTTGGTGCACCATATATGCTGTATATTT 60

Db 1 ATGAGTAAATATATATATATCGTGTCTGAATTTGGTGCACCATATATGCTGTATATTT 60

Thu May 13 11:53:12 2004

us-09-930-440b-7.rmp

Page 2

QY 61 GCAGAGAAATGATATTAAGCCAAAGAGCGCGGTATATGAGTAAATTCGAAACA 120
DB 61 GCAGAGAAATGATATTAAGCCAAAGAGCGCGGTATATGAGTAAATTCGAAACA 120
QY 121 TTTAAAGCTGATAAATTAATTCAGCTATTCGACCTAGGACAGATATCAATTAATAAAC 180
DB 121 TTTAAAGCTGATAAATTAATTCAGCTATTCGACCTAGGACAGATATCAATTAATAAAC 180
QY 181 ACAAGGAAATTAAGATCTCAGTTAGAAAGACAAATAAAGCTTGAATGAGATGACAT 240
DB 181 ACAAGGAAATTAAGATCTCAGTTAGAAAGACAAATAAAGCTTGAATGAGATGACAT 240
QY 241 TATTCATCATATGAGATTCAGTATCTTAAATTTGAGATGTTTCTACCCCTTTT 300
DB 241 TATTCATCATATGAGATTCAGTATCTTAAATTTGAGATGTTTCTACCCCTTTT 300
QY 301 GACGAGACTCTATTCATTTTTCAGTCTTTGAAACAAATAATGAGAAATCCCTTCA 360
DB 301 GACGAGACTCTATTCATTTTTCAGTCTTTGAAACAAATAATGAGAAATCCCTTCA 360
QY 361 GGTGAGTATGATTAATTCAGTATCTTGAATAATGACCAAGCTTCCGATCCGATGAG 420
DB 361 GGTGAGTATGATTAATTCAGTATCTTGAATAATGACCAAGCTTCCGATCCGATGAG 420
QY 421 AAATATATCATATCAACAGGAATGCTACTATTCATGATGATGATTAACAGTCTGTTTCTAT 480
DB 421 AAATATATCATATCAACAGGAATGCTACTATTCATGATGATGATTAACAGTCTGTTTCTAT 480
QY 481 TTTATATATTAATTAAGTTCGGTGTGTAATTAATTAATTAATTAATTAATTAATTAAT 540
DB 481 TTTATATATTAATTAAGTTCGGTGTGTAATTAATTAATTAATTAATTAATTAATTAAT 540
QY 541 CCAAGCGCTTGAAGATGTAACCTTAATGCTATTAATGATTTGAAATAACAATCCCT 600
DB 541 CCAAGCGCTTGAAGATGTAACCTTAATGCTATTAATGATTTGAAATAACAATCCCT 600
QY 601 AAGATATCAATGAGCTTCTCTGATCATTCAGCGGGTTTATGACATTTGCGCGGGT 660
DB 601 AAGATATCAATGAGCTTCTCTGATCATTCAGCGGGTTTATGACATTTGCGCGGGT 660
QY 661 CCTATGAGATTAATCTTTATTTGAAATAACAATTAATTAATTAATTAATTAATTAAT 720
DB 661 CCTATGAGATTAATCTTTATTTGAAATAACAATTAATTAATTAATTAATTAATTAAT 720
QY 721 GATCATTTGGCTCAATGAACTGATGAACTGAAACATCTTGTATGAGGATCAGTGT 780
DB 721 GATCATTTGGCTCAATGAACTGATGAACTGAAACATCTTGTATGAGGATCAGTGT 780
QY 781 GTTGAATAATCTTATGATTAATGATTAATGATTAATGATTAATGATTAATGATTAAT 840
DB 781 GTTGAATAATCTTATGATTAATGATTAATGATTAATGATTAATGATTAATGATTAAT 840
QY 841 ATCTGACGAGAAAGCTATTAATGCTTAAACAGATTAATAAAGGTGAGTGTTC 900
DB 841 ATCTGACGAGAAAGCTATTAATGCTTAAACAGATTAATAAAGGTGAGTGTTC 900
QY 901 GAAAAAATTAATTAACAAAGAAAGCTGATGATGATGATGATGATGATGATGATGAT 960
DB 901 GAAAAAATTAATTAACAAAGAAAGCTGATGATGATGATGATGATGATGATGATGAT 960
QY 961 TTTATGAGTAAATTTGAGAGCAAGACTTTATTCAGATGATTAATTAATTAATTAATTAAT 1020
DB 961 TTTATGAGTAAATTTGAGAGCAAGACTTTATTCAGATGATTAATTAATTAATTAATTAAT 1020
QY 1021 TTTCAAAATGAGGAGGAAATTAATGAGAAATAAATTAATTTG 1059
DB 1021 TTTCAAAATGAGGAGGAAATTAATGAGAAATAAATTAATTTG 1059

RESULT 2
US-09-870-759-83
Sequence 83, Application US/09870759
Patent No. US20020177551A1

GENERAL INFORMATION:
APPLICANT: TERMAN, David S
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT OF NEOPLASTIC DISEASE
FILE REFERENCE: 870759
CURRENT APPLICATION NUMBER: US/09/870,759
CURRENT FILING DATE: 2002-01-14
PRIOR APPLICATION NUMBER: US 60/208,128
PRIOR FILING DATE: 2000-05-30
NUMBER OF SEQ ID NOS: 166
SOFTWARE: PatentIn version 3.1
SEQ ID NO: 83
LENGTH: 17276
TYPE: DNA
ORGANISM: Streptococcus agalactiae
FEATURE:
NAME/KEY: CDS
LOCATION: (7062)..(8207)
OTHER INFORMATION:
US-09-870-759-83
Query Match 30.8%; Score 326.4; DB 9; Length 17276;
Best Local Similarity 58.3%; Pred. No. 1.7e-58;
Matches 617; Conservative 0; Mismatches 426; Indels 15; Gaps 2;
QY 1 ATGAGTAAATATATATATGCTGCTGAAATGAGTGAACCATATGATGATGATGATGAT 60
DB 12584 ATGAGTAAATATATATATGCTGCTGAAATGAGTGAACCATATGATGATGATGATGAT 12643
QY 61 GCAGAGAAATGATATTAAGCCAAAGAGCGCGGTATATGAGTAAATTCGAAACA 120
DB 12644 GCAGAGAAATGATATTAAGCCAAAGAGCGCGGTATATGAGTAAATTCGAAACA 12703
QY 121 TTTAAAGCTGATAAATTAATTCAGCTATTCGACCTAGGACAGATATCAATTAATAAAC 180
DB 12704 TTTAAAGCTGATAAATTAATTCAGCTATTCGACCTAGGACAGATATCAATTAATAAAC 12763
QY 181 ACAAGGAAATTAAGATCTCAGTTAGAAAGACAAATAAAGCTTGAATGAGATGACAT 240
DB 12764 ACAAGGAAATTAAGATCTCAGTTAGAAAGACAAATAAAGCTTGAATGAGATGACAT 12823
QY 241 TATTCATCATATGAGATTCAGTATCTTAAATTTGAGATGTTTCTACCCCTTTT 300
DB 12824 TATTCATCATATGAGATTCAGTATCTTAAATTTGAGATGTTTCTACCCCTTTT 12883
QY 301 GACGAGACTCTATTCATTTTTCAGTCTTTGAAACAAATAATGAGAAATCCCTTCA 360
DB 12884 GACGAGACTCTATTCATTTTTCAGTCTTTGAAACAAATAATGAGAAATCCCTTCA 12943
QY 361 GGTGAGTATGATTAATTCAGTATCTTGAATAATGACCAAGCTTCCGATCCGATGAG 420
DB 12944 GGTGAGTATGATTAATTCAGTATCTTGAATAATGACCAAGCTTCCGATCCGATGAG 12994
QY 421 AAATATATCATATCAACAGGAATGCTACTATTCATGATGATGATTAACAGTCTGTTTCTAT 480
DB 12995 AAATATATCATATCAACAGGAATGCTACTATTCATGATGATGATTAACAGTCTGTTTCTAT 13054
QY 481 TTTATATATTAATTAAGTTCGGTGTGTAATTAATTAATTAATTAATTAATTAATTAAT 540
DB 13055 TTTATATATTAATTAAGTTCGGTGTGTAATTAATTAATTAATTAATTAATTAATTAAT 13108
QY 541 CCAAGCGCTTGAAGATGTAACCTTAATGCTATTAATGATTTGAAATAACAATCCCT 600
DB 13109 CCAAGCGCTTGAAGATGTAACCTTAATGCTATTAATGATTTGAAATAACAATCCCT 13168
QY 601 AAGATTAATTAATTAAGTTCGGTGTGTAATTAATTAATTAATTAATTAATTAATTAAT 660
DB 13169 AAGATTAATTAATTAAGTTCGGTGTGTAATTAATTAATTAATTAATTAATTAATTAAT 13228
QY 661 CCTATGAGATTAATCTTTATTTGAAATAACAATTAATTAATTAATTAATTAATTAAT 720
DB 13229 CCTATGAGATTAATCTTTATTTGAAATAACAATTAATTAATTAATTAATTAATTAAT 13288
QY 721 GATCATTTGGCTCAATGAACTGATGAACTGAAACATCTTGTATGAGGATCAGTGT 780

DB 13289 GATGATTAAGCGCATGCAACACCGTATTTTACGCTTGTAGTTAAAGGGGTTCCGATT 13348
 QY 781 GTTGAAAAATCTTTAGGTTCAATAGTAAAGTGTACAGTTCCAGAAAGAAATTA 840
 DB 13349 GTTGCAACAGCGCTTAGGTGATTTGAAAAATCCAGATCCAGTGAAGAAAAATAG 13408
 QY 841 ATCTGACCAAGAAAGTCTATTATAGCTTAAACAGAGATTAATAAGGTGAGTTTTC 900
 DB 13409 ATTGTGCTGTATATCAGTGGTGTCTTTAAACCAATTAATAAAGGCAATTTATTC 13468
 QY 901 GAAAAAATATTAACAACAAAAGACCTGTATAGTATCAGTCCGATGAGAGTGTATAT 960
 DB 13469 ATAGAAAAATATAGCGGTGAACCGCCAGGTATGTGTCTCTATGAACGTGTATG 13528
 QY 961 TTATGGGTAAATATGACGACCAAGACTTATTCAGATGAATTAATATTCATAGCGAA 1020
 DB 13529 ATCTTGACCAAGAAAGCCAGATGATTTCCAGAGAGATGAATTTCTGTGATCAAG 13588
 QY 1021 TTCAAAAAATCAGGGGGAATTAATAGAACAAAAATTAAT 1058
 DB 13589 TTGAAATCAATTGCCCCGAGTATTAATACCACTTAAT 13626

RESULT 3
 US-09-751-708A-83

Sequence 83, Application US/09751708A
 Publication No. US20030157113A1
 GENERAL INFORMATION:
 APPLICANT: TERMAN, David S
 TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT OF NEOPLASTIC DISEASE
 FILE REFERENCE: 751708
 CURRENT APPLICATION NUMBER: US/09/751,708A
 CURRENT FILING DATE: 2002-10-15
 PRIOR APPLICATION NUMBER: US 60/173,371
 PRIOR FILING DATE: 1999-12-28
 NUMBER OF SEQ ID NOS: 166
 SOFTWARE: PatentIn version 3.1
 SEQ ID NO 83
 LENGTH: 17276
 TYPE: DNA
 ORGANISM: Streptococcus agalactiae
 FEATURE:
 NAME/KEY: CDS
 LOCATION: (7062)..(8207)
 OTHER INFORMATION:
 US-09-751-708A-83

Query Match 30.8%; Score 326.4; DB 10; Length 17276;
 Best Local Similarity 58.3%; Pred. No. 1.7e-58;
 Matches 617; Conservative 0; Mismatches 426; Indels 15; Gaps 2;

QY 1 ATGAGTAAATATATATGCTGCTGAATTTGTTCCAAACCTAATGCTAGTGTATAT 60
 DB 12584 AAGATTAAGCTTTATATATATGACAGATGCTGTGCAATCAATAGGAAATTAATCTT 12643
 QY 61 GCAAGAAATGATATTAAGCCCAAGAGCGGTGTTAATGACAGTAAATTCGAAACA 120
 DB 12644 GCGAAAAATATGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 12703
 QY 121 TTTAAGGTATTAATTAATTTGACCTTAATGCACTTAAGCAGAGTATCAATTAATAAC 180
 DB 12704 TTTAAGCTGAGAACTTATTTCTTAATTTGCTCCCAAGCTGAATATCAAAAGCACT 12763
 QY 181 ACAGAGAAATTAATGATTCAGTTAGAAATGACAAAAAGCTGAATAGATGAGAT 240
 DB 12764 ACAGAGAAATGAGATTAATTTGATGAGAGAAAGCTTTAGAGTTAGCTTTGAGAA 12823
 QY 241 TATTCATCTTAATGAAATATGCACTGATTAATTAATTAATGATTTTCTACCCCTTT 300
 DB 12824 TACTTAGAAATGCGATATATCAATTTCAAAAGGTGAGAACCTTTTCAACACCTTT 12883
 QY 301 GACGAAAGCTATTTGATTTTGTAGATTTTGAACAAAAATATGAGAAATCCCTTCA 360

DB 12884 GATGAAGAGCATATAGTTCCTTAATTTCTACAGATATGCCAATTTACAAATTCATCA 12943
 QY 361 GGTGAGTTATTAATTTACCGTATCTTGAAAAATATAGCCAGCTCCGATCCCTGATAG 420
 DB 12944 GAGAAATCACTATTTTACTTCTTATGAAAGATTTGGCAGAGAAC-----AAAG 12994
 QY 421 AAAATATCATATCAACAGAAATGCTACTATTTGATGAGATTAATAACAGTCTGTTCTAT 480
 DB 12995 AAGTATATCTTTCAGACGGGTATGCGGTATGGAAGAGATTCATCAAGCGGTGAATAT 13054
 QY 481 TTTAATAATTAATTAAGTTCGGTGTGAATATATACATATTTACATTTGCAATCTGATAT 540
 DB 13055 TTAGCGATGATGTCACACG-----ACATTTCTATTTTACATTTGTCACAGAGTAC 13108
 QY 541 CCAAGCCCTTTAGAGATGTAACCTTAATGCTATTAATGATTTGAAAAACCTTCCT 600
 DB 13109 CCAACACCTTACCCCTCTCTAATTTAAAGTATTAATCACTTTGAAAGATTAATTA 13168
 QY 601 AAGATTAATTAAGCTTTCTGATCACTTCTAGAGGGGTTATGACGATATGCGGCGGTG 660
 DB 13169 GATTTAAGATGATTTATGATGATCAATTTGATGATGATGATGATGATGATGATGAT 13228
 QY 661 CTTATGAAATTAACCTTTTATGAAAAACATTTCACTTTAGATTAATCTATGTCTGCGCA 720
 DB 13229 GCAATGGTGCAGAGATTAATTAATAAACCTTACTTATGATGATGATGATGATGAT 13288
 QY 721 GATCATTTGGCTTCAATGAACTGATGAACTGAAACCTTTGATTTGGGGTCAAGTGT 780
 DB 13289 GATCATTAACCGCATGTCACACCTGATTTTAACTGCTGCTTATGATTAAGGGTGTGCAAT 13348
 QY 781 GTTGAAAAATCTTTAGGTTCAATATAGTAAAGTGTACAGCTTCAGAAAGAAATTA 840
 DB 13349 GTTGCAACAGCGCTTAGGTGATTTGAAAAATCCAGATCCAGTGAAGAAAAATAG 13408
 QY 841 ATCTGACCAAGAAAGTCTATTATAGCTTAAACAGAGATTAATAAGGTGAGTTTTC 900
 DB 13409 ATTGTGCTGTATATCAGTGGTGTCTTTAAACCAATTAATAAAGGCAATTTATTC 13468
 QY 901 GAAAAAATTAACAACAAAAGACCTGTATAGTATCAGTCCGATGAGAGTGTATAT 960
 DB 13469 ATAGAAAAATATAGCGGTGAACCGCCAGGTATGTGTCTCTATGAACGTGTATG 13528
 QY 961 TTATGGGTAAATATGACGACCAAGACTTATTCAGATGAATTAATATTCATAGCGAA 1020
 DB 13529 ATCTTGACCAAGAAAGCCAGATGATTTCCAGAGAGATGAATTTCTGTGATCAAG 13588
 QY 1021 TTCAAAAAATCAGGGGGAATTAATAGAACAAAAATTAAT 1058
 DB 13589 TTGAAATCAATTGCCCCGAGTATTAATACCACTTAAT 13626

RESULT 4

US-10-192-280-1
 Sequence 1, Application US/10192280
 Publication No. US2004009574A1
 GENERAL INFORMATION:
 APPLICANT: Datagupta, Nandhubhan
 TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING
 TITLE OF INVENTION: STREPTOCOCCUS AGALACTIAE CAPSULAR POLYSACCHARIDE SYNTHESIS
 TITLE OF INVENTION: GENES
 FILE REFERENCE: 475412001300
 CURRENT APPLICATION NUMBER: US/10/192,280
 CURRENT FILING DATE: 2002-07-09
 NUMBER OF SEQ ID NOS: 39
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 1
 LENGTH: 25020
 TYPE: DNA
 ORGANISM: Streptococcus agalactiae
 US-10-192-280-1

Thu May 13 11:53:12 2004

us-09-930-440b-7.rnpb

Page 4

Query Match	30.7%	Score	324.8	DB	16	Length	25020
Best Local Similarity	58.2%	Pred. No.	4.3e-58				
Matches	616	Conservative	0	Mismatches	427	Indels	15
						Gaps	2

QY	1	ATGAGCAATATATATATATCGTTCGTGAATATGGTCTCAACCAATATGAGTATGATCATTT	60
Db	13086	AAGATATATCGTTATATATATATTCAGAGANTGGTTGCATCATATATGAGATATTTAATCTT	13145
QY	61	GCAAGCAAAATGATATTTAAAGCCAAAGAGCCCGGTATATATGCAATATATCCAAACA	120
Db	13146	GCGAAAAAATATGATATATGTTCCGGTGTCTTGTGGTGAAGCTTTAAATTTTCAGACT	13205
QY	121	TTTAAAGCTGATTAATTTATTTTCACTATTTGACCTAAGGCGAGATATCAATATAAAAC	180
Db	13206	TTTAAAGCTGAGAACTTATTTTCTAAATTTGGTCCCAAGCTGATATCAAAAAGAACT	13265
QY	181	ACAGAGAAATTAGATCTCAGTTAGAAATGACAAAAAAGCTTGAATATGATATGACAT	240
Db	13266	ACAGGAACAGCAACAGTCACACTTGGATGAGAACGAAACGTTTAGAGTTAAGCTTTAAAGAA	13325
QY	241	TATCTCCATCTAATGGAATATGACATCAAGTTAAATTTAGATGTTTTTCTACCCCTTT	300
Db	13326	TACTTGAATATGCGTATATATGCAATTTCAAAAGGTGGAACCTTTTCAACACTTTT	13385
QY	301	GACCAAGACTATATGATTTTTTTCATCTTTGAAACAAAAATATGAAAAATCCCTTCA	360
Db	13386	GATGAAGAGCATATAGAGTTCTTAATTTCTACAGATATGCCAATTTACAAAATTCATCA	13445
QY	361	GGTAGATTAATGAATTTACCGTATCTTGAAAAAAATGACCAAGCTTCGATCCCTGATAG	420
Db	13446	GGAGAAATCACTAATTTACTCTTCTAGAAAAATTTGGCAAGAAC-----AAAAG	13496
QY	421	AAAAATATCATATCAACAGSATACTACTATGATGAGATTAACCAAGTCTGTTCTAT	480
Db	13497	AAAGTATTCCTTCGACGGGTATGCGGTATGGAAGAGATCATCATACGCTGATATTT	13556
QY	481	TTTATATAATATTAAGTCCGGTTGGAAATATATACATATATACATGCAATCTGATAT	540
Db	13557	TTTACGTACATATGATAC-----TGACATTTCTATTTTACTTTGTCACACAGATAC	13610
QY	541	CCAAAGCCCTTTGAGAGATGAACCTTAATGCTAATTAATGATTTGAAAAACACTCCCT	600
Db	13611	CCAAAGCCTTACCCCTCTCTAAATTTAAACGTTATTCATACTTTGAAGATGAATTTAA	13670
QY	601	AAGATTAACATAGAGCTCTCTGATCAATCTAGCGGGTTTATACAGCTATTCGCGCGTG	660
Db	13671	GATTTAACGATAGGTTTATGGATCATTTGATGATGATGATGATGATGATGATGATGATG	13730
QY	661	CGTTATGAATTAATTTATGAAAAACATTTCACTTATGATTAATTCATATGCTGSCCA	720
Db	13731	GCAATGGGTGCAGAGTTATGAAAAACATTTTACTTATGATCTAATATGAAAGTCCG	13790
QY	721	GATCATTTGGCTCAATAGAACCTGATGAACGTAACATCTTGTATTTGGGCTCAGTCT	780
Db	13791	GATATTAAGCCAGTGCACACCTGATTTTATTTTACGCTCTTTATTTAAAGGGTTTGCAT	13850
QY	781	GTTGAAAAATCTTTAGGTTCAAAATGTAAGTGGTTACAGCTTCAGAAAGAAATTA	840
Db	13851	GTTGAAACAAGCTTAGGTATTTGAAAAATCCAGATCCAGTATGAAAGAAAAATTAAG	13910
QY	841	ATCGTACCAAGAAAGTGTATTTATAGCTAAACAGAGATTAATAAAAGGTGAGTTTTTCA	900
Db	13911	ATTGTTGCTGTTAATCAGTCGTTCTTTAAACCAATTTAAAGGCCGATATTTATCA	13970
QY	901	GAATAAAATATTAACAACAAAAGCTGGTATATGATCAGTCCGATGAGTGTATAT	960
Db	13971	ATTAACAAATATTAAGGTGAAGCCGCAAGTAAATGTTCTCCTTAAGACGTATGAC	14030
QY	961	TTATTTGGTAAATTTGACAGACAGACTTATTCAGATGAATTAATTAATTCATAGCAA	1020
Db	14031	ATCTTTGGACAAAGAACCCAGATGATTTCCAAAGGATGAAGTATTCCTGATTCACGC	14090
QY	1021	TTCAAAATCAAGGAGATTAATGAGCAAAATTTATTT 1058	

Db 14091 TTGAAATCAATGCCCGAGTTATAATTAACACCTAAT 14128

```

RESULT 5
US-09-816-028A-1
: Sequence 1, Application US/09816028A
: Patent No. US20020042369A1
: GENERAL INFORMATION:
: APPLICANT: Gilbert, Michel
: APPLICANT: Makarchuk, Warren W.
: APPLICANT: National Research Council of Canada
: TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
: TITLE OF INVENTION: Gangliosides and Ganglioside Mimics
: FILE REFERENCE: 019633-00011US
: CURRENT APPLICATION NUMBER: US/09/816, 028A
: PRIOR FILING DATE: 2001-03-21
: PRIOR APPLICATION NUMBER: US 60/118,213
: PRIOR FILING DATE: 1999-02-01
: PRIOR APPLICATION NUMBER: US 09/495,406
: PRIOR FILING DATE: 2000-01-31
: NUMBER OF SEQ ID NOS: 49
: SOFTWARE: Sequlin Ver. 2.1
: SEQ ID NO 1
: LENGTH: 11474
: TYPE: DNA
: ORGANISM: Campylobacter jejuni
: FEATURE:
: OTHER INFORMATION: 11.5 kb PCR product from C. jejuni OH4384
: OTHER INFORMATION: including LOS biosynthesis locus
US-09-816-028A-1

```

Query Match	Similarity	9.9%	Score 105.2	DB 9	Length 11474	
Best Local	Similarity	49.1%	Pred. 8.5e-12			
Match	479	Conservative	0	Mismatches 456	Indels 39	Gaps 6
Qy	17	TCGTTGCTGMAATTGGTTGCAACCATTAATGGTAGTGTGATATTCGAAAGAGAAATGATAT	76			
Db	6979	TCGTGCTCGAATATGGCATTTATCATTAATGCGAGTTTGAACCTAAATTAATGATAG	7038			
Qy	77	TAAAGCCAAAGAGCCCGGTGTTATGCAATTAATTCGAAACATTAAAGCTGATTAAT	136			
Db	7039	ATGCGACCTTTAGCAGAGTGTGAAGATTATTAACATCAAAACCAATCGTTTGAAGATG	7099			
Qy	137	TAAATTCAGTATTCGACCTAAGCAGAGTATCAATATAAAAACAAGAGATTAAGAT	196			
Db	7099	AGATAG-----TAAGCCGCTAAAAAGTATTCGTGATATGCAAAATTA	7144			
Qy	197	CTCAGTTAGAAATGCAAAAAAGCTTAAATGAGATGACATTATCTCCATCTAATGG	256			
Db	7147	GCAATTAAGAGTTATGCAAAATGCGCTTAGATTATTAAGATGAGCTAGACCTTAAG	7200			
Qy	257	AATATGCGCTCGTTAAATTTAGATGTTTTTTTACCCCTTTGACGAAAGACTATAG	316			
Db	7207	AATACACGAAATTAAGTCTCTGTTTATCTTAGCACACTTTTCTCGTCAGGTGCAA	7266			
Qy	317	ATTTTTTAGCATCTTGAACCAAAAAATATGAAAAATCCCTTCAAGTGAATTAATG	376			
Db	7267	ACCGCTTGAAGATATGGAGTTAGCTTTTAAATATGGTTCAAGGTGAGTTAAATAT	7322			
Qy	377	TACCGTATCTTAAAAAATATGGCAAGCTTCGATCCCGATATAGAAAAATATCATATCA	436			
Db	7327	ATCCGCTATTAAACACATAG-----CAGCCTTTAAABAGCTATGATATGTTAGCA	737			
Qy	437	CAGGAATGCGCTACTATGATGAGATTAACAAGCTGTGTTCTATTTTATTAATATTAAG	496			
Db	7378	CAGGATGATATGTTTGAAGTATTAACATTAACCAAGTAAAAAATCTTATATAGCAATGAAA	7433			
Qy	497	TTCCGGTGTATATATACATATTAACATTGCAATATCTGAATATCCAAAGCCCTTGAAG	556			
Db	7438	TTCCCTTTGTT-----TTAAATGACACAAACCATTTTAAACCAACCCGACATATTC	748			
Qy	557	ATGTAAAGCTTAATGCTATTAATGATTTAAAAAACAATCCCTTAAGATTAACATAGCT	616			

```

Db 7489 TTGTAAATTAAGCGCTTACCTGATTAATAAAGAAATTTCT---TGATGTAGGCT 7545
Qy 617 TCTCTGATCACTTACCGGCTTTTATGACCTATTCGCGGCGCTTATGAAATTAATT 676
Db 7546 TAAGCGCACACAAACAGATATCTGCGTATGAGGTGCGCTTGGCTTGTG 7605
Qy 677 TTAATGAAAACATTCACCTTACATTAATCTATGCTGGGCCAGATCATTTGGCCCAA 736
Db 7606 TCGCTGAAGACATTTTACTGATAGTATGCAATGAAGTGGCCCTGATATAGTTGTTCTA 7665
Qy 737 TAGAACCTGATGAACCTGAACATCTTGTATGATGGGCTCAAGTGTGTGAAAAATCTTAG 796
Db 7666 TGGATACACAGCGCTTAAAGAGCTTATPATACAAAGTGAGCAAAATGGCTATATGAGAG 7725
Qy 797 GTTCAATATGTAAGCTTACAGCTTACAGAAAGGAAGATTAATGTAGCAAGAAAGT 856
Db 7726 GAAATATATGAAGTAAAAAGAGAGCTAAGCAAGTCAATGATTTTGCCTTGG 7785
Qy 857 CTATTATAGCTTAAAC---AGAGTAAAAAAGGTGAGCTTTTTCAGAAAAAATATAA 933
Db 7786 CAAGCGTACAGCATTAAGATATTTAAAAAGCGAAGTTTATCTATGATTAATTTT 7845
Qy 914 CAACAAAAGACCTGG---TAATGCTATGATCCGATGAGCTGATTAATTTATGGCTA 970
Db 7846 GGGTTAAAAAGCCTGAGCTTGATGATGATGAGCTGATTAATTTATTTAGCA 7905
Qy 971 AAATGACAGACAGA 986
Db 7906 AAAAGCATTAAGAGA 7921

RESULT 6
US-10-303-161-1
; Sequence 1, Application US/10303161
; Publication No. US20030148459A1
; GENERAL INFORMATION:
; APPLICANT: Gilbert, Michel
; APPLICANT: Makarchuk, Warren W.
; APPLICANT: National Research Council of Canada
; TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
; TITLE OF INVENTION: Campylobacter and Ganglioside Mimics
; FILE REFERENCE: 019633-000111US
; CURRENT APPLICATION NUMBER: US/10/303,161
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US/09/816,028
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/118,213
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: US 09/495,406
; PRIOR FILING DATE: 2000-01-31
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 11474
; TYPE: DNA
; ORGANISM: Campylobacter jejuni
; FEATURE:
; OTHER INFORMATION: 11.5 kb PCR product from C. jejuni OH4384
; OTHER INFORMATION: Including LOS biosynthesis locus
; US-10-303-161-1

```

Query Match 9.9%; Score 105.2; DB 15; Length 11474;
 Best local Similarity 49.1%; Pred. No. 8,5e-12;
 Matches 479; Conservative 0; Mismatches 458; Indels 39; Gaps 6;

```

Qy 17 TCGTTGCTGAATTTGTTGCAACCATATGCTAGTGTGATTTATGCAAGAAATGATAT 76
Db 6979 TCGGCTGCTGAATTTGCGCATTAATCATATGCGAGTTTGAACATGATTAATTAATGATAG 7038
Qy 77 TAAAGCCAAAGAGCGCGGTGTTAATGACGTAATAATTCAAACATTTAAAGCTGATTAAT 136
Db 7039 ATGAGCGCTTTAGACACAGGTGCTAATGATTTAAGCATCAAAACCCACATGCTTGAAGATG 7098

```

```

Qy 137 TAATTCAGCTATTCGACCTAAGGACAGATATCAATTAATAAAGACAGAGGATTAAGAT 196
Db 7099 AGATGAG-----TAGGCGCTTAATAAAGTAAATTTCTGATATGCAAAATATA 7146
Qy 197 CTCAGTTAGAAATGACAAAAGGCTTGAATGAGATGACGATTAATTCATTAATG 256
Db 7147 GCATTTATGAGATTAAGCAAAAATGCTTATGATTAATAAGATGAGCTAGACCTTAAG 7206
Qy 257 AATATGACGCTTAAATTAATTAAGTGTTTTCTATCCCTTTGAGCAAGATCATTTG 316
Db 7207 AATACACAGAAAATTAAGGCTCTGTTATCTTACACACCTTTTCTCGTGCAGGTGCAA 7266
Qy 317 AATTTTACCATCTTGAACAAATAATGAAAAATCCCTTCAAGGTGATTAATTAAT 376
Db 7267 ACCGCTTAAGAAATAGGAGATGATGCTTTTAAGATGTGTCAGGTGAGTAAATTAAT 7326
Qy 377 TACCGTATCTGAAAAATAGCCAGCTTCGATCCCTGATAGAAATATCATATCA 436
Db 7327 ATCCGCTTATTAACACATAG-----CAGCTTAAAGCCTATGATATGTTAGCA 7377
Qy 437 CAGGATGCTACTATGATGAGATTAACAGCTGCTTTCTATTTTATTAATTAATTAAG 496
Db 7378 CAGGATGATATGATTAAGATTAACAGATTAACAGATTAACAGATTAATTAAG 7437
Qy 497 TTCCGTTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 556
Db 7438 TTCCCTTGT-----TTAATGACACACACCATTTTAAACCCAGCCGATTAATC 7488
Qy 557 ATGTAACCTTAAGCTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 516
Db 7489 TTGTATGATTAACGCTATGCTGATTAATTAATTAATTAATTAATTAATTAATTAAT 7545
Qy 617 TCTCTGATCACTTACCGGCTTTTATGACCTATTCGCGGCGCTTATGAAATTAATT 676
Db 7546 TAAGCACACACACAGATTAATCTTGGCTGTTAGGTGCTGATGCACTTGGCTGCTG 7605
Qy 677 TATTTGAAAACATTCACCTTACATTAATTCATGCTGCGGCGGCTTATGAGCTCAA 736
Db 7606 TCGTTAAAAAGCATTTTACTGATAGTATGATGATGAAGTGGCCCTGATATAGTTGTTCTA 7665
Qy 737 TAGAACCTGATGAACCTGAAACATCTTGTATGATGGGCTCAAGTGTGTGAAAAATCTTAG 796
Db 7666 TGGATACACAGCGCTTAAAGAGCTTATPATACAAAGTGAGCAAAATGGCTATATGAGAG 7725
Qy 797 GTTCAATATGTAAGCTTACAGCTTACAGAAAGGAAGATTAATGTAGCAAGAAAGT 856
Db 7726 GAAATATATGAAGTAAAAAGAGAGCTAAGCAAGTCAATGATTTTGCCTTGG 7785
Qy 857 CTATTATAGCTTAAAC---AGAGTAAAAAAGGTGAGCTTTTTCAGAAAAAATATAA 913
Db 7786 CAAGCGTACAGCATTAAGATTAATAAAGGAGAAATTTATCTATGATTAATTTT 7845
Qy 914 CAACAAAAGACCTGG---TAATGCTATGATCCGATGAGCTGATTAATTTATGGCTA 970
Db 7846 GGGTTAAAAAGCCTGAGCTTGATGATGATGAGCTGATTAATTTATTTAGCA 7905
Qy 971 AAATGACAGACAGA 986
Db 7906 AAAAGCATTAAGAGA 7921

```

RESULT 7
 US-10-303-118-1
 ; Sequence 1, Application US/10303118
 ; Publication No. US20030157655A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Gilbert, Michel
 ; APPLICANT: Makarchuk, Warren W.
 ; APPLICANT: National Research Council of Canada
 ; TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
 ; TITLE OF INVENTION: Campylobacter and Ganglioside Mimics
 ; FILE REFERENCE: 019633-000111US

Thu May 13 11:53:12 2004

us-09-930-440b-7.rmpb

Page 6

CURRENT APPLICATION NUMBER: US/10/303,118
CURRENT FILING DATE: 2002-11-21
PRIOR APPLICATION NUMBER: US/09/816,028
PRIOR FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: US 60/118,213
PRIOR FILING DATE: 1999-02-01
PRIOR APPLICATION NUMBER: US 09/495,406
PRIOR FILING DATE: 2000-01-31
NUMBER OF SEQ ID NOS: 49
SOFTWARE: Patent Ver. 2.1
SEQ ID NO 1
LENGTH: 11474
TYPE: DNA
ORGANISM: Campylobacter jejuni
FEATURE:
OTHER INFORMATION: 11.5 kb PCR product from C. jejuni OH4384
OTHER INFORMATION: including LOS biosynthesis locus
US-10-303-118-1

Query Match 9.9%; Score 105.2; DB 15; Length 11474;
Best Local Similarity 49.1%; Pred. No. 8.5e-12;
Matches 479; Conservative 0; Mismatches 458; Indels 39; Gaps 6;

17 TCGTGTGCTGAATGTTGGTGAACATATAGTGTGTTGATTTGCAAGAGAAATATAT 76
DB 6979 TCGTGTGCTGAATGTTGGTGAACATATAGTGTGTTGATTTGCAAGAGAAATATAT 7038
QY 77 TAAAGCCAAAGAGCCGGTGTATATGACATTAATTCACACATTTAAAGCTGATTAAT 136
DB 7039 ATGACGCTTTAGACAGAGTCTTAAGTTAAAGATCAACACCATGTTGAAGATG 7098
QY 137 TAAATTAGCTATGCACTTAAAGCAAGTATCAAAATTAATAAACAAGAGAAATTAAT 136
DB 7099 AGATGAG-----TAAAGCCCTTAATAAAGTATCTGTAATGCAAAATAA 7146
QY 197 CTCAGTAAAGTAAATGACAAATAAGCTTAATAATGAGATTAATTCATCTCAATTAATG 256
DB 7147 GCATTTAGATATGCAAAATATGCTTTAGTTAAAGATGAGTGAAGCTTAAG 7206
QY 257 AATATGACATGCTTAATTAATTAAGATTTTCTTACCCCTTTGACAAAGACTTAATG 316
DB 7207 AATACAGAGAAATTAAGTCTTGTATCTTACCAACCTTTCTGTCGACAGTCCAA 7266
QY 317 ATTATTAGCATCTTTGAAACAAATAATGAGAAATCCCTGAGTGAATTAATTAAT 376
DB 7267 ACCGCTTAGAAGATATGAGATTAAGTCTTTAGTTAAAGTGTTCAGTGAAGTGAAT 7326
QY 377 TACCGTATCTTGAATAAATAGCCAAAGCTCCGATCCCTGATTAAGAAATTAATCAAT 436
DB 7327 ATCCGCTTAATTAACATAG-----CAGCTTTAAAGAGCTATATAGTTAGCA 7377
QY 437 CAGGAATGCTATGATGATGAGTAAACAGTGTGTTTATATTTTAAATAATTAAG 496
DB 7378 CAGGAGTAATGATTAAGTAAAGTAAACCACTGTAATAATTTATTAAGCAAGAA 7437
QY 497 TTCGGTGTGTAATTAATTAACATTAATGCAATGCAATTAATTAACAGCCCTTGAAG 556
DB 7438 TTCCTTTGTT-----TTAATGCAACAACCATCTTTACCCCAACCCGATATATC 7488
QY 557 ATGTAACCTTAATGCTTAATTAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 616
DB 7489 TGTGAATTAATTAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 7545
QY 617 TCTGTGATCTATTTAGAGGGGTTTATGACGCTATGCGGCGTGTCTTATGATTAAT 676
DB 7546 TAAAGCAACCAACAAGATTAATCTGCGTGTGTTAGTGGGTTGACCTTGCGTTG 7605
QY 677 TTAATGAAAACATTTCACTTAATTAATTAATTAATTAATTAATTAATTAATTAAT 736
DB 7606 TGGCTGAAAGACATTTTATGATTAATTAATTAATTAATTAATTAATTAATTAAT 7665
QY 737 TAAAGCTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 796

DB 7666 TGAATACACAGCTTTAAAGAGCTTAATTAATTAATTAATTAATTAATTAATTAAT 7725
QY 797 GTTCAATAGTAAATGCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 856
DB 7726 GAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 7785
QY 857 CTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 913
DB 7786 CAAGCTAGTCACTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 7845
QY 914 CAACAAAGAGCTG-----TAAATGATTAATTAATTAATTAATTAATTAATTAATTAAT 970
DB 7846 GGGTTAAAGACCTGATGCTGTTGTAATTAATTAATTAATTAATTAATTAATTAATTAAT 7905
QY 971 AATTCAGAGCAAGA 986
DB 7906 AAAAGCATTAAGAG 7921

RESULT 8

US-10-303-128-1
Sequence 1, Application US/10303128
Publication No. US2003015765A1
GENERAL INFORMATION:
APPLICANT: Gilbert, Michel
APPLICANT: Makarchuk, Warren W.
TITLE OF INVENTION: National Research Council of Canada
TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
FILE REFERENCE: 019633-0001105
CURRENT APPLICATION NUMBER: US/10/303,128
CURRENT FILING DATE: 2002-11-21
PRIOR APPLICATION NUMBER: US/09/816,028
PRIOR FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: US 60/118,213
PRIOR FILING DATE: 1999-02-01
PRIOR APPLICATION NUMBER: US 09/495,406
PRIOR FILING DATE: 2000-01-31
NUMBER OF SEQ ID NOS: 49
SOFTWARE: Patent Ver. 2.1
SEQ ID NO 1
LENGTH: 11474
TYPE: DNA
ORGANISM: Campylobacter jejuni
FEATURE:
OTHER INFORMATION: 11.5 kb PCR product from C. jejuni OH4384
OTHER INFORMATION: including LOS biosynthesis locus
US-10-303-128-1

Query Match 9.9%; Score 105.2; DB 15; Length 11474;
Best Local Similarity 49.1%; Pred. No. 8.5e-12;
Matches 479; Conservative 0; Mismatches 458; Indels 39; Gaps 6;

17 TCGTGTGCTGAATGTTGGTGAACATATAGTGTGTTGATTTGCAAGAGAAATATAT 76
DB 6979 TCGTGTGCTGAATGTTGGTGAACATATAGTGTGTTGATTTGCAAGAGAAATATAT 7038
QY 77 TAAAGCCAAAGAGCCGGTGTATATGACATTAATTCACACATTTAAAGCTGATTAAT 136
DB 7039 ATGACGCTTTAGACAGAGTCTTAAGTTAAAGATCAACACCATGTTGAAGATG 7098
QY 137 TAAATTAGCTATGCACTTAAATTAATTAAGTAAATTAATTAATTAATTAATTAAT 196
DB 7099 AGATGAG-----TAAAGCCCTTAATAAAGTATCTGTAATGCAAAATAA 7146
QY 197 CTCAGTAAAGTAAATGACAAATAAGCTTAATAATGAGATTAATTCATCTCAATTAATG 256
DB 7147 GCATTTAGATATGCAAAATATGCTTTAGTTAAAGATGAGTGAAGCTTAAG 7206
QY 257 AATATGACATGCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 316
DB 7207 AATACAGAGAAATTAAGTCTTGTATCTTACCAACCTTTCTGTCGACAGTCCAA 7266

QY	317	ATTTTTGCACCTTTGAAACAAAATAATGGAATAATCCCTCAGGTGATTAATTAAT	376
Db	7267	ACCGTTGAAAGATATGGAAGTAACTGTTTAAATGCTTCAGGTGATTAATTAAT	7322
QY	377	TACCGTATCTTGAAAAAATPAGCAAGCTTCGATCCCTGATTAAGAAAAATATCATATCA	436
Db	7327	ATCCGCTATTAAACACATAG-----CAGCCTTTAAAAAGCCTATGATAGTTAGCA	7377
QY	437	CAGGAATGCGTACTATATGATAGATAAAACAGCTCGTTTCTATTTTAAATATAATAG	496
Db	7378	CAGGATTAATATGATTTGAAAGTATTAACCTGTTAAAAATCTTATTAAGCAATGAAA	7433
QY	497	TTCCGGTGTGATATATACAAATATTACATGTCATATGAAATATCCAAAGCCCTTGAAG	556
Db	7438	TTCCCTTTGTT-----TTAATGCAACAACCAATCTTATACCAACCCGCAATATC	7488
QY	557	ANGTAAACCTTAATGCTATTAATGATTTGAAAAACCTCCCTGGAATATAATAGCT	616
Db	7489	TTGTAAGATTAAACGCTAATGCTTAATTAATAAAAGAAATTTCT--TGATGTAGGCT	7544
QY	617	TCTCGATCATTTCTAGCGGGTTTATGAGACTATTCGCGCGGTGCTTATGSAATPACT	676
Db	7546	TAAGGACCAACACACAGATATCTTGCGTGTTAGTGCGGTTGCACTTGGTGCCTG	7605
QY	677	TTATTGAAAAACATTTCACCTTATAGATAATCTATGCTTGCCCAATCATTTGGCCTCA	736
Db	7606	TGCTGAAAGCAATTTACTCTAGTATGATGCAATAGAGTGGCCCTGATATAGTTGTCTA	7665
QY	737	TAGAACCTGATGAACCTGAACACATCTTTATTTGGGGTCAAGGTGTGTAATAATCTTAG	796
Db	7666	TGAGTACACAGGCTTAAAGAGCTTATTATACAAAGTAGCAATGCTATTAATAGAG	7722
QY	797	GTTCAATATGTAAGTGTATCAGCTTCAGAAAGAGATTAATCTAGCAAGAAAGT	856
Db	7726	GAATTAATGAAGATTAATAAAGAGAGGTAGCAGAGAACAGTCAATTAATTTTGCTTGG	7785
QY	857	CTATTATGCTAAAC--AGAGATPAATAAAGAGTGAAGTATTTTCAGAAAAAATATTA	913
Db	7786	CAAGGTATGTCAGCTTAAGATATTAATAAAGGGAAGTTTATCTATGAGAAATATTT	7844
QY	914	CAACAAAAGACCTGG--TAATGTAATGCTCCATGAGTGAAGTGTATTAATTTAGGCTA	970
Db	7846	GGGTTAAGACCTGAGCTGTGTAATTTAGTCAGAGCTGAATTTGAAATATTTATTTAGCA	7905
QY	971	AAATGTCAGACAAG 986	
Db	7906	AAAAAGCATTAAGGA 7921	

RESULT 9
 US-10-303-134-1
 Sequence 1, Application US/10303134
 Publication No. US20030157657A1
 GENERAL INFORMATION:
 APPLICANT: Gilbert, Michel
 APPLICANT: Makarchuk, Warren W.
 APPLICANT: National Research Council of Canada
 TITLE OR INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
 TITLE OR INVENTION: Gangliosides and Ganglioside Mimics
 FILE REFERENCE: 019633-00011US
 CURRENT APPLICATION NUMBER: US/10/303,134
 CURRENT FILING DATE: 2002-11-21
 PRIOR APPLICATION NUMBER: US/09/826,028
 PRIOR FILING DATE: 2001-03-21
 PRIOR APPLICATION NUMBER: US 60/118,213
 PRIOR FILING DATE: 1999-02-01
 PRIOR APPLICATION NUMBER: US 09/495,406
 PRIOR FILING DATE: 2000-01-31
 NUMBER OF SEQ ID NOS: 49
 SOFTWARE: PatentIn Ver. 2.1
 SEQ ID NO 1
 LENGTH: 11474
 TYPE: DNA

```

; ORGANISM: Campylobacter jejuni
;
; FEATURE:
; OTHER INFORMATION: 11.5 kb PCR product from C. jejuni OH438
; OTHER INFORMATION: including LOS biosynthesis locus
US-10-303-134-1

```

Query Match	9.9%;	Score 105.2;	DB 15;	Length 11474;
Best Local Similarity	49.1%;	Pred. No. 8.5e-12;		
Matches 479;	Conservative 0;	Mismatches 458;	Indels 39;	Gaps 6

QY	17	TCGTCGCAAAATGCTGCAACACATATGATGCTGATATGCAAGACAAATGAT	76
Db	6979	TCGTCGCTGAATGCGCATATATCATATGCGACGTTTGAACTGCTAAATATATGCTG	703
QY	77	TAAAGCCAAAGAGCCGCGTGTAAATGCAATTAATTCGAACATTTAAAGCTGATAT	136
QY	137	TAAATTGACGATATGCACCTAAGCAGAGTATCAATTAATAAACAAGAGATTAAGAT	196
Db	7039	ATGCAAGCCCTTAGCACAAGTGCCTAAGATTAATAAGCATCAACCCACATCTGTGAAGATG	7091
QY	7099	AGATGAG-----TAGGCCGTAAATAAGTAAATTCCTGTATGCAAAATTA	714
QY	197	CTCAGTTAGAAATGACAAAAGCCTTGAAATGAAGTATGACGATATCTCATATAG	256
Db	7147	GCATTATATGAGATTATGCAAAAATGTCGCTTAGATTATAAAGATGAGCTAGACTTAAG	7200
QY	257	AATATGACGTCAGTTAAATTTAATATGTTTTCTACCCCTTACAGACGACTATATG	316
Db	7207	AATACACGAAAAATAGTCTGTGTATCTTAGCACACTTTTCTCTGCGAGGTGCA	7266
QY	317	ATTATTAGCATCTTGTGAACAAAATAATGAAAATTCCTTCAGGTAGTATTAAT	376
Db	7267	ACCGCTTAGAAGATATGCGATGTGTGCTTTTAAGATGTGTCAGGTAGCTAATAT	7322
QY	377	TACCGTATCTGAAAAAATAGCCAGCTCCGATCCCTGATAGAAAATATCATATCA	436
Db	7327	ATCCGCTATATTAACACATAG-----CAGCCTTAAAAAGCCTATGATAGTTAGCA	7377
QY	437	CAGGAATGCTACTATGATGATGATTAACACAGTCTGTTCATTTTATAATATAAG	496
Db	7378	CAGGATGAAATGATATGAAAGTATTAACCACTGTAATAATCTTATAGCATGAA	743
QY	497	TTCCGCTGTGTAATATTAACAATTTACATGCAATCTGAAATATCAAGCCCTTTGAG	556
Db	7438	TTCCCTTTGTT-----TTAATGCAACAACAACATTTTACCAACCCCGCATATTC	7488
QY	557	ATGTAACCTTAATGCTATTAAATGATTTGAATAAACACTCCCTAGATATAATAGGCT	616
Db	7489	TTGTAAGATTAAAGCTATGCTTAATTAATAAATAAGATTTTCT---TGATATGATGCT	7544
QY	617	TCTCTGATCATCTTAGCGGGTTTATGCAAGCTATTCGCGCGTGCCTTATGGAATACTT	676
Db	7546	TAAACGACACACACAACAGATATCTTGCGTGTATTAGGTCGCGTGCCTTATG	7606
QY	677	TTATTTGAAAAACATTCACCTTAATAAATCATATGTCGGCCAGATCATTTGGCCTCA	736
Db	7606	TGCTTGAAGACATTTTATCTGATGATGACATAGAGTGGCCCTGATATAGTTGTCTTA	7666
QY	737	TAGAAGCTGATGAACAGACATCTTTGATATGCGGTCAGGTGTGTTGAAAAATTTTAG	796
Db	7666	TGATATACACAGGCTTTAAAGAGCTTATTTATACAAGAGACAAATGCTATTAATGAG	7722
QY	797	GTTCAATATGTAAGTGTTCACCTTCAGAAAGGAAGAATTAATTCGTAGCAAGAAGT	856
Db	7726	GAAATATATGAAGTAAATAAGACACTAGCAAGAGCAAGTCAATTAATTTTGCTTTG	7788
QY	857	CTATTATAGCTAAATAC---AGAGATTAATAAATAAGTGAAGTTTTTTCAGAAAAAATATTA	913
Db	7786	CAGCGTATGTCAGATTAAAGATTTAAAAAAGGGAAGTTTATCTATGATATATTT	7844
QY	914	CACAAAAAAGACTCG---TAATGATACATCCGATGAGAGTGTGATATTTATTTGGGTA	970
Db	7846	GCGTTAAAGACTGAGCTGTGTAATTTAGTCAAGCTAAATTTGAAATATTTTAAAGCA	7900

Thu May 13 11:53:12 2004

us-09-930-440b-7.rnpb

Page 8

QY 971 AATTGACAGACAGA 986
DB 7906 AAAAGCATTAGAGA 7921

RESULT 10

US-10-303-162-1
Sequence 1, Application US/10303162
Publication No. US20030157658A1
GENERAL INFORMATION:
APPLICANT: Gilbert, Michel
APPLICANT: Maxarichuk, Warren W.
APPLICANT: National Research Council of Canada
TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
TITLE OF INVENTION: Gangliosides and Ganglioside Mimics
FILE REFERENCE: 019633-000111US
CURRENT APPLICATION NUMBER: US/10/303,162
CURRENT FILING DATE: 2002-11-21
PRIOR APPLICATION NUMBER: US/09/816,028
PRIOR FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: US 60/118,213
PRIOR FILING DATE: 1999-02-01
PRIOR APPLICATION NUMBER: US 09/495,406
PRIOR FILING DATE: 2000-01-31
NUMBER OF SEQ ID NOS: 49
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 11474
TYPE: DNA
ORGANISM: Campylobacter jejuni
FEATURES:
OTHER INFORMATION: 11.5 kb PCR product from C. jejuni OH4384
OTHER INFORMATION: including LOS biosynthesis locus
US-10-303-162-1

Query Match 9.9%; Score 105.2; DB 15; Length 11474;
Best Local Similarity 49.1%; Pred. No. 8.5e-12;
Matches 479; Conservative 0; Mismatches 458; Indels 39; Gaps 6;

QY 17 TCGTGTCTGAATGTGTGACACATATAGTGTATGTATGCAAGAGAAATGAT 76
DB 6379 TCGTGTCTGAATGTGTGACATATATATATAGTGTATGCAAGAGAAATGAT 7038
QY 77 TAAAGCCAAAGAGCCGGGTGTATGATGAGTAAATCCAAATTTAAAGTATAT 136
DB 7039 ATGACGCTTTAGACAGAGTCTTATGATATTAAGCATCAACCAATGCTTGAAGT 7098
QY 137 TAAITTCAGCTATGACCTTAAGCAGATATCAATTAATAAACAAGGAATTAAT 196
DB 7099 AGATGAG-----TAAAGCCGCTAAATAAGTATTCCTGTATATGCAAAATAA 7146
QY 197 CTCGATAGAAATGACAAAGAGTTGAATGAATGAGATGAGATATCTCATATAG 256
DB 7147 GCATTTATGATATGCAAAAGTGTCTTTAGATTTAAAGATGAGCTTGAAG 7206
QY 257 AATAAGAGTCAGTTTAAATTAATTAATGATGTTTCTACCCCTTTAGCAAGACTATG 316
DB 7207 AATAAGAGTAAATTAATGATGTTTCTTATCTTACACACCTTTTCTGAGAGTCA 7266
QY 317 ATTTTATGATCTTTGAAACAAATAATGAAATATCCCTGAGGATATTAAT 376
DB 7267 ACCGCTTAAAGATATGAGATGATGCTTTTAAAGATGCTTGAAGATTAAT 7326
QY 377 TACCGTATCTTGAATAATAGCAAGCTTCGATCCCTGATAGAAATATATCAATCA 436
DB 7327 ATCCGCTATTAACACATAG-----CAGCTTTAAAGACCTATGATATTAAGCA 7377
QY 437 CAGAAATGCTATCTTATGATGATTAATAGAGTCTGTTTCTATTTATTAATTAAG 496
DB 7378 CAGGAATGATATGATTAATGAATTAATTAATTAATTAATTAATTAATTAAT 7437
QY 497 TCCGCTTGTATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 556

DB 7438 TCCCTTTGTT-----TTAATGACACACACCAATCTTTTACCACCCGCAATATAC 7488
QY 557 ATGTAACTTATATGCTATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 616
DB 7489 TGTATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 7545
QY 617 TCTGTATCATCTTACGCGGTGTTTATGACGCTATGCGCGGTGCTTATGAATACT 676
DB 7546 TAAGCACCACACACACATTAATCTGCTGTTTATGAGTGGGTGCTTGGCTTGG 7605
QY 677 TATGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 736
DB 7606 TGTGTAAGACATTTTACTGATATGATGATGATGATGATGATGATGATGATG 7665
QY 737 TAGAACCTGATGAACCTGAACATCTTGTATGCGGTGCTGATGATGATGATGAT 796
DB 7666 TGGATACACAGGCTTAAAGAGCTTATATACAAAGTGAACAAATGCTATATAGAG 7725
QY 797 GTTCAATATGAATAGGCTTACAGCTTACAGAAAGAAATTAATGATAGCAAGAAAT 856
DB 7726 GAAATATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAAT 7785
QY 857 CATTATAGCTTAAAC--AGAGATTAATAAGTGAAGTGTGTTTTCAGAAATAATTA 913
DB 7786 CAGGCTAGTACGATTAATGAATTAATTAATTAATTAATTAATTAATTAATTA 7845
QY 914 CAAATAAGAGCTG--TATGTATGATGATGATGATGATGATGATGATGATGAT 970
DB 7846 GGGTAAAGAGCTGATGCTTGTGGAATTAATGACCTGAATTTGAATAATTTAGGA 7905
QY 971 AATTGACAGACAGA 986
DB 7906 AAAAGCATTAGAGA 7921

RESULT 11

US-09-882-227-489
Sequence 489, Application US/09882227
Publication No. US20030158396A1
GENERAL INFORMATION:
APPLICANT: Kleanchous, Harold
APPLICANT: Al-Garawi, Amal
APPLICANT: Miller, Charles
APPLICANT: Tomb, Jean-Francois
APPLICANT: Oomen, Raymond P.
TITLE OF INVENTION: Identification of Polynucleotides
TITLE OF INVENTION: Encoding No. US20030158396A1 Helicobacter polypeptides in th
FILE REFERENCE: 06132/047002
CURRENT APPLICATION NUMBER: US/09/882,227
CURRENT FILING DATE: 2001-06-15
PRIOR APPLICATION NUMBER: US 08/902,615
PRIOR FILING DATE: 1997-07-29
NUMBER OF SEQ ID NOS: 638
SOFTWARE: SeqSIO for Windows Version 4.0
SEQ ID NO 489
LENGTH: 1123
TYPE: DNA
ORGANISM: Helicobacter pylori
FEATURES:
NAME/KEY: CDS
LOCATION: (19)...(1038)
US-09-882-227-489

Query Match 6.6%; Score 70.4; DB 10; Length 1123;
Best Local Similarity 44.4%; Pred. No. 6.6e-05;
Matches 434; Conservative 0; Mismatches 526; Indels 18; Gaps 3;

QY 14 ATATCTGCTGTAATATGTTGATGACATTAATGATGATGATGATGATGATGATGAT 73
DB 35 AATATGCTGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 94

Thu May 13 11:53:12 2004

us-09-930-440b-7.rnpb

Page 9

QY 74 TATTAAAGCCAAAGAGCGGCTGTATATGAGTAATTCCTAAACATTAAAGCTATA 133
DB 95 TTCAAGCCATTAAAGAGCGGCTGTATATGAGTAATTCCTAAACATTAAAGCTATA 154
QY 134 AATTAAATTCAGCTATTCAGCTTAAGAGAGTATCAATTAATAAACAAGAGATTTAG 193
DB 155 GCATGACTTTAACTCTAAAGAGATTCCTTCATCATTCAGAGCACTTTATAGGATAAG 214
QY 194 AATTCAGTTAGAAATGACAAAAGCTTGAATATAGATATGACATTAATCTCCATCTAA 253
DB 215 AAAATTTATGAAATGTATCAAAAAGGCTTTACCCCCCTAGAAATGACATCTGATTTGT 274
QY 254 TGAATATGAGCTAGCTTTAAATTTAGATGTTTTTCTACCCCTTTTACAGAGCTCTA 313
DB 275 TTGAGTTGGCTTAAAGAGCTGATTTAGGCAATTTTACGCTTTTATGATCAGAGCTT 334
QY 314 TTGATTTTTCAGATCTTTGAAAACAAAATATGAAAATCTCTTACAGTGAATTTGA 373
DB 335 TAGAGCTTTTGAAGAGCTTAATTCCTCATATATAAATGCTAGTTTGAATATCTTG 394
QY 374 AATTACCGTATCTGAAAATATGCGCAAGCTCCGATCCCTGATAGAAAATTAATCATAT 433
DB 395 AATTGAGCTTATTAATAAGAGCGGCTCG-----CACAAAAGCCCATTAATCTTT 445
QY 434 CAACAGAAATGCTACTATTTGATGATTAACAGCTCTGTTCTATTTTAAATAATA 493
DB 446 CTAGAGGATATGCTACACACAGGATGCAAGAGCTATCTCATTTGAGAGAG----- 500
QY 484 AAGTCCCGGTTGATATTAAGAAATTCATGCAATTCGAAATATCCAGCCCTTTG 553
DB 501 -AGTAATATTTTACATCACCTTTTAAATGCTGAGCGCTTATCCAGTAATAATAG 559
QY 554 AGGATGTAACCTTATCTATTTATGATTTGAAAACAACTCTCCCTAAGATTAACATAG 613
DB 560 AAGAGCTTAATCTTGAAGATGTTAAATTTAGGGAATCTT---TGGCTTAAATTTG 616
QY 614 GCTTCTGATCATTTCTAGCGGCTTTATGACGATTTGCGGCGGCTTATGAAATA 673
DB 617 GCTTGAAGATCAACGATTTGCTCTTTGCCCCATTTTACCCCACTTTAGAGAGCA 676
QY 674 CTTTATGAAAACATTTCACTTTAGATTAATCTATGCTGCGCCAGATCATTTGAGCT 733
DB 677 GCATATATGAAAAGATTTCAATTTAAACAAATCTTAAACCCAGAGCGCTTTTA 736
QY 734 CAATGAACTGATGAACTGAAACATCTTTGATTTGGGCTGAGTGTGTTGAAAATCT 793
DB 737 GCATGATTTTAAACGATTTTAAACAGATTTGAAGCAATCAAGCAAGCGCTTTAGCT 796
QY 794 TAGCTTCAATATGTAAGTGTATACGCTTCGAAGAGAAATAAATTCGACAGAA 853
DB 797 TAGGCAAGAGAGCGCAAGATCAATCCAAAGACTTTTGAAGAGGAAATTTTTCGAC 856
QY 854 ACTTATATATAGCTAAACAGATTAATAAAGTGAAGTCTTTTTCAGAAAAAATATPA 913
DB 857 GCTCTTTATTTGTATTAAGATATCAAAAAGGAGAGCATGACTGAAAAAATATCA 916
QY 914 CAACAAAAGAGCTGATATGATATGATGAGCGGATGAGAGTGAATTTATGGGTAATA 973
DB 917 AAGCCTTAGCGCCCAACTGCTTACACCTTAATTTTATTAAGAAAATTTTAGCCAAA 976
QY 974 TTGAGAGAGAGACTTTA 991
DB 977 AAGCATCAAAATTTCTAA 994

RESULT 12
US-09-984-205-3
Sequence 3, Application US/09984205
Patent No. US20020137175A1
GENERAL INFORMATION:
APPLICANT: Coleman, Timothy A. et al.
TITLE OF INVENTION: Human Glycosylation Enzymes
FILE REFERENCE: PFS05D1

CURRENT APPLICATION NUMBER: US/09/984,205
CURRENT FILING DATE: 2001-10-29
PRIOR APPLICATION NUMBER: PCT/US00/05325
PRIOR FILING DATE: 2000-03-01
PRIOR APPLICATION NUMBER: 09/516,143
PRIOR FILING DATE: 2000-03-01
PRIOR APPLICATION NUMBER: 60/122,409
PRIOR FILING DATE: 1999-03-02
NUMBER OF SEQ ID NOS: 6
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 3
LENGTH: 1080
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(1077)
US-09-984-205-3
Query Match 5.9%; Score 63; DB 9; Length 1080;
Best Local Similarity 47.8%; Pred. No. 0.0024;
Matches 183; Conservative 0; Mismatches 200; Indels 0; Gaps 0;
QY 518 TATTACATTCGATATGCAATATTCACAGCCCTTTGAGAGTGAACCTTAATGCTATA 577
DB 542 TCTTGAGATGATACAGGCGCATACCGCTCCAGCTTGAGAGCTACACCTGGGCTCATCT 601
QY 578 ATGATTTGAAAACAACTCTCCCTAAGATTAACATAGCTTCTGATCATTTAGCGGCT 637
DB 602 CGGAATATGCAAGCTCTTTCTGATCATTCATAGGATTTCTGGCATTAACAGCA 661
QY 638 TTATGAGCTATTTGCGGCGGCTCTATGGAATTAATTTATGAAAACATTTCACTT 697
DB 662 TAGGATATCTGTCGCGCAGATGCTCTGCGGCGCAAGGTTGAAAGCTACATACTT 721
QY 698 TAGTAATCTATGCTGCGCCAGATCATTTGCGCTCAATAGAACTGATGAACTGAAAC 757
DB 722 TGAAGAGCTGAGAGGAGATGACCACTGCGCTGCTGAGACCTGAGAACTGGCG 781
QY 758 ATCTTTGATTTGGGCTGAGCTGTGTTGAAAATCTTTAGCTCAATATGTAAGTGTTA 817
DB 782 AGCTGTGCGGCTCAGCTCTGCTGTGAGAGCTGCGCTCCCAACAGAGAGCTGC 841
QY 818 CAGCTTCAAGAGGAAATTAATCTGAGCAAGAACTATATAGCTTAACAGAGA 877
DB 842 TGCCCTGTGAGATGCTCTGCAATGAGAAAGCTGGGCAAGCTGTGTGTGCTGCAAGTAAA 901
QY 878 TAAAAAAGGTGAGGTTTTTTCA 900
DB 902 TTCGGAAGGACCATTTCTAACA 924

RESULT 13
US-09-930-440B-5
Sequence 5, Application US/09930440B
Patent No. US20020142386A1
GENERAL INFORMATION:
APPLICANT: Aetnaugh et al.
TITLE OF INVENTION: Engineering Intracellular Signaling Pathways
FILE REFERENCE: PFS09P2
CURRENT APPLICATION NUMBER: US/09/930,440B
CURRENT FILING DATE: 2001-08-16
PRIOR APPLICATION NUMBER: 60/227,579
PRIOR FILING DATE: 2000-08-25
PRIOR APPLICATION NUMBER: 09/516,793
PRIOR FILING DATE: 2000-03-01
PRIOR APPLICATION NUMBER: 60/169,624
PRIOR FILING DATE: 1999-12-08
PRIOR APPLICATION NUMBER: 60/122,582
PRIOR FILING DATE: 1999-03-02
NUMBER OF SEQ ID NOS: 8
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 5

LENGTH: 1080
 TYPE: DNA
 ORGANISM: Homo sapiens
 FEATURE:
 NAME/KEY: CDS
 LOCATION: (1)..(1080)
 US-09-930-440b-5

Query Match 5.9%; Score 63; DB %; Length 1080;
 Best Local Similarity 47.8%; Pred. No. 0.0024;
 Matches 183; Conservative 0; Mismatches 200; Indels 0; Gaps 0;

QY 518 TATTACATTGCANATCTGAATATCCAAAGCCCTTTGAGATGTAAACCTTAATGCTATTA 577
 Db 542 TCTTCAGATGTACACAGGCAATACCCGCTCCAGCCTGAGACGTCACCTGCGGGTCACT 601
 QY 578 ATGATTTGAAAAACAATTCCTTACAGATTAATAGAGCTTCTGATCAATTTAGAGGGT 637
 Db 602 CGGAATATCGAAGAGCTTTCTCTGACATTCCTCAATAGGATTTCTGGGCAATGAAACAGCA 661
 QY 638 TTATGCACTATTGCGCGGCTCTTATGGAATAACTTTATGAAAAACATTTCACTT 697
 Db 662 TAGGATATCTGTGGCGGAGTGGCTCTGGGGCCAAAGGTGTGAAACGTCAATTAATT 721
 QY 698 TAGATTAATCTATGCTGTGGCCCAATCATTTGGCTCAATAGAACGTGATGAACTGAAAC 757
 Db 722 TGAACAAGACCTTGAGAGGAGATGACCACTGGCTGTGGAGCTTGAGAACTGAGCCG 781
 QY 758 ATCTTTATTTGGGGTCAGGTGTGTGAAAAATCTTTAGTTCAAAATAGTAAGGTGTA 817
 Db 782 AGCTGGTGGGTCAGTGCCTGTGTGAGAGGTGCGCTGCCCAACCAACAGCACTGC 841
 QY 818 CAGCTTCAGAAAGAGAAATTAATGCTAGCAAGAAAGTATATTAATGCTAAACAGAGA 877
 Db 842 TGCCTGTGAGATGGCTCTGCAATGAGAAAGCTGGGCAAGTGTGGTGGCCAAAGTGA 901
 QY 878 TAAAAAAGGTGAGGTTTTTTCA 900
 Db 902 TTCGGAAGGACCACTTCTAAC 924

RESULT 14
 US-10-037-270-1035
 Sequence 1035, Application US/10037270
 Publication No. US20030104529A1
 GENERAL INFORMATION:
 APPLICANT: Tang, Y. Tom
 APPLICANT: Liu, Chenghua
 APPLICANT: Asundi, Vinod
 APPLICANT: Zhang, Jie
 APPLICANT: Ren, Feiyan
 APPLICANT: Chen, Rui-hong
 APPLICANT: Zhao, Qing A.
 APPLICANT: Weinman, Tom
 APPLICANT: Xue, Aidong J.
 APPLICANT: Yang, Yonghong
 APPLICANT: Wang, Jian-Rui
 APPLICANT: Zhou, Ping
 APPLICANT: Ma, Yungding
 APPLICANT: Wang, Dunhui
 APPLICANT: Wang, Zhiwei
 APPLICANT: Tillinghast, John
 APPLICANT: Drmanac, Radoje T.
 TITLE OF INVENTION: No. US20030104529A1 Nucleic Acids and
 FILE REFERENCE: Polypeptides
 CURRENT APPLICATION NUMBER: US/10/037,270
 PRIOR FILING DATE: 2002-01-04
 PRIOR APPLICATION NUMBER: 09/552,317
 PRIOR FILING DATE: 2000-04-25
 PRIOR APPLICATION NUMBER: 09/488,725
 PRIOR FILING DATE: 2000-01-21
 NUMBER OF SEQ ID NOS: 1104

SOFTWARE: pc FL genes Version 1.0
 SEQ ID NO 1035
 LENGTH: 1230
 TYPE: DNA
 ORGANISM: Homo sapiens
 FEATURE:
 NAME/KEY: CDS
 LOCATION: (119)..(1198)
 US-10-037-270-1035

Query Match 5.9%; Score 63; DB 15; Length 1230;
 Best Local Similarity 47.8%; Pred. No. 0.0025;
 Matches 183; Conservative 0; Mismatches 200; Indels 0; Gaps 0;

QY 518 TATTACATTGCANATCTGAATATCCAAAGCCCTTTGAGATGTAAACCTTAATGCTATTA 577
 Db 660 TCTTCAGATGTACACAGGCAATACCCGCTCCAGCCTGAGACGTCACCTGCGGGTCACT 719
 QY 578 ATGATTTGAAAAACAATTCCTTACAGATTAATAGAGCTTCTGATCAATTTAGAGGGT 637
 Db 720 CGGAATATCGAAGAGCTTTCTCTGACATTCCTCAATAGGATTTCTGGGCAATGAAACAGCA 779
 QY 638 TTATGCACTATTGCGCGGCTCTTATGGAATAACTTTATGAAAAACATTTCACTT 697
 Db 780 TAGGATATCTGTGGCGGAGTGGCTCTGGGGCCAAAGGTGTGAAACGTCAATTAATT 839
 QY 698 TAGATTAATCTATGCTGTGGCCCAATCATTTGGCTCAATAGAACGTGATGAACTGAAAC 757
 Db 840 TGAACAAGACCTTGAGAGGAGATGACCACTGGCTGTGGAGCTTGAGAACTGAGCCG 899
 QY 758 ATCTTTATTTGGGGTCAGGTGTGTGAAAAATCTTTAGTTCAAAATAGTAAGGTGTA 817
 Db 900 AGCTGGTGGGTCAGTGCCTGTGTGAGAGGTGCGCTGCCCAACCAACAGCACTGC 959
 QY 818 CAGCTTCAGAAAGAGAAATTAATGCTAGCAAGAAAGTATATTAATGCTAAACAGAGA 877
 Db 960 TGCCTGTGAGATGGCTCTGCAATGAGAAAGCTGGGCAAGTGTGTGGTGGCCAAAGTGA 1019
 QY 878 TAAAAAAGGTGAGGTTTTTTCA 900
 Db 1020 TTCGGAAGGACCACTTCTAAC 1042

RESULT 15
 US-10-117-722-1035
 Sequence 1035, Application US/10117722
 Publication No. US20030219744A1
 GENERAL INFORMATION:
 APPLICANT: Tang, Y. Tom
 APPLICANT: Liu, Chenghua
 APPLICANT: Asundi, Vinod
 APPLICANT: Zhang, Jie
 APPLICANT: Drmanac, Radoje T.
 TITLE OF INVENTION: No. US20030219744A1 Nucleic Acids and
 FILE REFERENCE: Polypeptides
 CURRENT APPLICATION NUMBER: US/10/117,722
 PRIOR FILING DATE: 2002-04-04
 PRIOR APPLICATION NUMBER: 09/620,312
 PRIOR FILING DATE: 2000-07-19
 PRIOR APPLICATION NUMBER: 09/552,317
 PRIOR FILING DATE: 2000-04-25
 PRIOR APPLICATION NUMBER: 09/488,725
 PRIOR FILING DATE: 2000-01-21
 NUMBER OF SEQ ID NOS: 1104
 SOFTWARE: pc FL genes Version 1.0
 SEQ ID NO 1035
 LENGTH: 1230
 TYPE: DNA
 ORGANISM: Homo sapiens
 FEATURE:
 NAME/KEY: CDS
 LOCATION: (119)..(1198)

US-10-117-722-1035

Query Match 5.9%; Score 63; DB 16; Length 1230;
Best Local Similarity 47.8%; Pred. No. 0.0025;
Matches 183; Conservative 0; Mismatches 200; Indels 0; Gaps 0;

QY	518	TATTACATTCGAAATCTGAATATCCAGCCCTTGAGAGATTAACCTTAATGCTATTA	577
DB	660	TCCTTCAGCTGACCGGCAATCCCGCTCCAGCCTGAGAGCTCACTGCGGCTCATCT	719
QY	578	ATGATTTAAAAACACTTCCTTAAGAAATACATAGCCTTCTGATCATTTAGCGGCT	637
DB	720	CGAATATACAGAGCTCTTCCCTGACATTCCTATAGGGTATTTGCGCATGAAACAGGCA	779
QY	638	TTTATGAGCTATGCGCGGCTGCTTAAGAAATTAATTTATGAAAAACATTTCACTT	697
DB	780	TAGCATATCTGTGGCCGAGTGCTCTGCGGCGCAAGTGTTGAACTGACATATCTT	839
QY	698	TAGATTAATCTATGCTGCGCCCAATCATTTGGCCTCAATAGAACCTGATGAACTGAAC	757
DB	840	TGACAAAGACCTGGAAGGGAGTGAACACTCGGCTCGCTGGAAGCTGAGAACTGCGCG	899
QY	758	ATCTTTGATTTGGGCTCAGGCTGTGTAATAATCTTTAGCTTCAATAGTAAAGTGTTA	817
DB	900	AGCTGTGCGGCTCAGTGCGTCTTGAGAGCGTGCCCTGGGCTCCCAACCAAGCAGCTGC	959
QY	818	CAGCTTCAGAAAGAGAAATAAATCTAGCAAGAAAGCTATTTAGCTAAACAGAGA	877
DB	960	TGCCCTGTGAGTGGCTGCAATGAGAAAGCTGGGCAAGTCTGTGTGTCGCAAGTGAATA	1019
QY	878	TAAAAAAGGTGAGGTTTTTCA	900
DB	1020	TTCCGAGAGCACCATTTCTAACA	1042

Search completed: May 12, 2004, 12:37:38
Job time : 501.708 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 08:55:49 ; Search time 48.7787 seconds
(without alignments)
2004.184 Million cell updates/sec

Title: US-09-930-440B-8

Perfect score: 1778

Sequence: 1 MSNIYIVAEICMNGSVDI.....RQDFLPPELLHSEFNKGE 346

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

A_Geneseq_29Jan04:*
1: Geneseqp1980s:*
2: Geneseqp1990s:*
3: Geneseqp2000s:*
4: Geneseqp2001s:*
5: Geneseqp2002s:*
6: Geneseqp2003as:*
7: Geneseqp2003bs:*
8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1778	100.0	346	AAV96102	AAV96102 Escherich
2	1778	100.0	346	AAV96102	AAV96102 Escherich
3	1778	100.0	346	AAV96102	AAV96102 Escherich
4	1016.5	57.2	338	AAV96102	AAV96102 Escherich
5	979.5	55.1	341	AAV96102	AAV96102 Escherich
6	516.5	29.0	359	AAV96102	AAV96102 Escherich
7	514.5	28.9	359	AAV96102	AAV96102 Escherich
8	514.5	28.9	359	AAV96102	AAV96102 Escherich
9	510.5	28.7	359	AAV96102	AAV96102 Escherich
10	510.5	28.7	359	AAV96102	AAV96102 Escherich
11	510.5	28.7	359	AAV96102	AAV96102 Escherich
12	510.5	28.7	359	AAV96102	AAV96102 Escherich
13	497.5	28.0	349	AAV96102	AAV96102 Escherich
14	491.5	27.6	400	AAV96102	AAV96102 Escherich
15	475.5	26.7	346	AAV96102	AAV96102 Escherich
16	475.5	26.7	346	AAV96102	AAV96102 Escherich
17	449.5	25.3	328	AAV96102	AAV96102 Escherich
18	402	22.6	340	AAV96102	AAV96102 Escherich
19	389	21.9	338	AAV96102	AAV96102 Escherich
20	238.5	13.4	201	AAV96102	AAV96102 Escherich
21	224.5	12.6	182	AAV96102	AAV96102 Escherich
22	174.5	9.8	123	AAV96102	AAV96102 Escherich
23	114	6.4	500	AAV96102	AAV96102 Escherich
24	112.5	6.3	1365	AAV96102	AAV96102 Escherich
25	110.5	6.2	1365	AAV96102	AAV96102 Escherich

26	110	6.2	769	6	ABR51668	ABR51668 Protein s
27	109	6.1	601	6	ABU27219	ABU27219 Protein e
28	108.5	6.1	363	6	ABM70260	ABM70260 Photornab
29	107.5	6.0	713	6	ABR53234	ABR53234 Protein s
30	104.5	5.9	436	6	AAV95776	AAV95776 Human pro
31	103.5	5.8	623	6	ABR53138	ABR53138 Protein s
32	103	5.8	653	4	AAV82533	AAV82533 Human hea
33	103	5.8	653	4	AAV12985	AAV12985 Human HSP
34	103	5.8	653	4	AAV33206	AAV33206 Human mit
35	103	5.8	950	6	ABU19361	ABU19361 Protein e
36	102	5.7	481	5	ABP73535	ABP73535 Candida a
37	102	5.7	1241	6	ABU24033	ABU24033 Protein e
38	101.5	5.7	650	5	ABP26071	ABP26071 Streptoco
39	101.5	5.7	650	6	ABU46537	ABU46537 Protein e
40	101.5	5.7	841	3	AAV05948	AAV05948 Protein d
41	101	5.7	125	4	ABG17251	ABG17251 Novel hum
42	101	5.7	455	4	ABU52748	ABU52748 Human nuc
43	101	5.7	455	5	AAV47363	AAV47363 Human DEA
44	101	5.7	455	5	AAE20272	AAE20272 Human lun
45	101	5.7	455	6	ADA54078	ADA54078 Human pro

ALIGNMENTS

RESULT 1
AAV96102
ID AAV96102 standard; protein; 346 AA.

XX AAV96102;
XX 19-DEC-2000 (first entry)
XX Escherichia coli stialic acid synthetase.
XX DE Escherichia coli stialic acid synthetase.
XX KM Stialic acid synthetase; human; sas gene; sialylation; glycoprotein;
XX KM plasmidogen; transferin; chylotrocin; Na⁺,K⁺-ATPase.
XX OS Escherichia coli.
XX PN MO200052135-A2.
XX PD 08-SEP-2000.
XX PF 01-MAR-2000; 2000WC-US005313.
XX PR 02-MAR-1999; 99US-0122582P.
XX PR 08-DEC-1999; 99US-0169624P.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PA (UTCO) UNIV JOHNS HOPKINS.
XX PA (UYMY-) UNIV WYOMING.
XX PI Betsenbaugh MJ, Lawrence S, Lee YC, Jarvis D, Coleman TA;
XX WPI: 2000-572178/53.
XX DR N-PSDB; AAV50570.
XX PT Recombinant production of sialylated glycoproteins using cells in which
XX PT the expression of enzymes, e.g. stialic acid synthetase, involved in the
XX PT sialylation reaction has been altered.
XX PS Example 6; Page 108; 144pp; English.
XX CC The present sequence is that of Escherichia coli stialic acid synthetase,
XX CC encoded by the neuB gene (see AAV50570). The neuB gene was used to
XX CC isolate the corresponding human gene (see AAV50569) for stialic acid
XX CC synthetase (see AAV96101). The invention provides methods and
XX CC recombinantly engineered cells for producing glycoproteins having
XX CC sialylated oligosaccharides. The methods involve altering the expression
XX CC of enzymes involved in carbohydrate processing. A claimed cell producing
XX CC sialylated glycoprotein at above endogenous levels expresses at least 1
XX CC (preferably human) enzyme selected from GlcNAc-2 epimerase, an enzyme

XX Vaccine; glycoprotein; insect cell; enzyme; N-acetylglucosamine-2;
 KW GLNAC-2; epimerase; UDP-GlcNAc; mannose; (Man)NAC; stialic acid;
 KW synthetase; aldolase; cytidine monophosphate-stialic acid; CMP-SA;
 KW transporter; stialylated glycoprotein.
 XX
 XX Escherichia coli.
 XX
 XX US2002142386-A1.
 XX
 XX 03-OCT-2002.
 XX
 XX 16-AUG-2001; 2001US-00930440.
 XX
 XX 02-MAR-1999; 99US-0122582P.
 XX
 XX 08-DEC-1999; 99US-0169624P.
 XX
 XX 25-AUG-2000; 2000US-0227579P.
 XX
 XX (BETE/) BETENAUICH M J.
 XX (LAWR/) LAWRENCE S.
 XX (LEEY/) LEE Y C.
 XX (COLE/) COLEMAN T A.
 XX
 XX Betenbaugh MJ, Lawrence S, Lee YC, Coleman TA;
 XX
 XX WPI: 2003-102519/09.
 XX
 XX N-PSDB; AA153994.
 XX
 XX Manipulating glycoprotein production in insect cell, involves enhancing
 PT expression of enzymes involved in carbohydrate processing pathway such as
 PT N-acetylglucosamine-2 epimerase or stialic acid synthetase.
 XX
 XX Disclosure; Fig 35D; 88pp; English.
 XX
 XX The invention relates to a novel method for manipulating glycoprotein
 CC production in an insect cell comprising enhancing expression of an
 CC enzyme, such as N-acetylglucosamine-2 (GlcNAc-2) epimerase, one
 CC catalysing conversion of UDP-GlcNAc to mannose (Man)NAC, stialic acid
 CC synthetase, aldolase, cytidine monophosphate-stialic acid (CMP-SA)
 CC synthetase or CMP-SA transporter, where the expression of each enzyme is
 CC enhanced to above endogenous levels. The novel method is useful for
 CC manipulating glycoprotein production in an insect cell. Further methods
 CC of the invention are useful for producing stialylated glycoprotein. The
 CC stialylated glycoprotein produced by the above mentioned methods are
 CC useful as pharmaceutical compositions, vaccines, diagnostics and
 CC therapeutics. This sequence represents the bacterial stialic acid (NeuB)
 CC synthetase protein of the invention
 CC
 XX Sequence 346 AA;
 SQ

Query Match 100.0%; Score 1778; DB 6; Length 346;
 Best Local Similarity 100.0%; Pred. No. 4.8e-163;
 Matches 346; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSNIIYVAEICGNHNSVDIAEMILAKKAGVNAVFPQTKAKLISALPKAEYQIKN 60
 DB 1 MSNIIYVAEICGNHNSVDIAEMILAKKAGVNAVFPQTKAKLISALPKAEYQIKN 60
 QY 1 TGELESQLEMTKKLEMKYDDYILHMEYAVSLNLDVFPSTPEDSIDFLASIKOKIMWIPS 120
 DB 1 TGELESQLEMTKKLEMKYDDYILHMEYAVSLNLDVFPSTPEDSIDFLASIKOKIMWIPS 120
 QY 121 GELINLPYLEKIAKLPIDPKKIIISTGMAITIDEIKOSYSIFINNKVPGNTIIHCTEY 180
 DB 121 GELINLPYLEKIAKLPIDPKKIIISTGMAITIDEIKOSYSIFINNKVPGNTIIHCTEY 180
 QY 181 PTPEDVNAIINLKKHFPKONIGFSDHSGFAAALAAVYGTIPTEKHTTLKSGSGP 240
 DB 181 PTPEDVNAIINLKKHFPKONIGFSDHSGFAAALAAVYGTIPTEKHTTLKSGSGP 240
 QY 241 DHLASIBEDDELKHLICIGVRCVEKISGNSKVVTASERKNKIIVAKSIIIAKTEIKGGEVFS 300
 DB 241 DHLASIBEDDELKHLICIGVRCVEKISGNSKVVTASERKNKIIVAKSIIIAKTEIKGGEVFS 300

QY 301 EKNIITRPNGISPMEMVNLKIAEQDEIPDELIHSEPKNOGE 346
 DB 301 EKNIITRPNGISPMEMVNLKIAEQDEIPDELIHSEPKNOGE 346

RESULT 4
 ID AAY68965 standard; protein; 338 AA.
 XX
 XX AAY68965;
 AC
 XX
 DT 30-MAY-2000 (first entry)
 XX
 XX Cps2P protein which is involved in stialic acid synthesis.

XX Capsular gene cluster; serotype 2; polysaccharide biosynthesis;
 KW capsular component; antigen; regulation; chain length determination;
 KW complement-mediated opsonophagocytosis; serotype-specific detection;
 KW antigen; vaccine; Streptococcal disease; ORF 22; ORF 21; ORF 22; Cps2A;
 KW Cps2B; Cps2C; Cps2D; Cps2E; Cps2F; Cps2G; Cps2H; Cps2I; Cps2J; Cps2K;
 KW Cps2O; Cps2P; Cps2Q; Cps2R; Cps2S; Cps2T.
 XX
 XX Streptococcus suis.
 OS
 XX
 XX NO200005378-A2.
 XX
 XX 03-FEB-2000.
 PD
 XX
 XX 19-JUL-1999; 99NO-NL000460.
 PF
 XX
 XX 22-JUL-1998; 98EP-00202465.
 PR 22-JUL-1998; 98EP-00202467.
 XX
 XX (DIEN-) STICHTING DIENST LANDBOUWKUNDIG ONDERZOE.

XX Smith HE;
 XX
 XX WPI: 2000-195104/17.
 DR N-PSDB; AA260929.
 DR
 XX
 XX New nucleic acid containing the capsular gene cluster of Streptococcus
 PT suis, used for serotype-specific detection and to generate antigens or
 PT mutants for vaccination.
 XX
 XX Disclosure; Fig 3; 144pp; English.
 PS

XX The proteins AAY68950-69 are encoded by the capsular gene cluster of
 CC Streptococcus suis serotype 2. The genes in this cluster are involved in
 CC polysaccharide biosynthesis of capsular components and antigens. The
 CC proteins are involved in regulation (CpsA), chain length determination
 CC (CpsB, CpsC), export (CpsC'), and biosynthesis (CpsR, CpsF, CpsG, CpsH,
 CC CpsJ, CpsK). The capsule confers bacterium resistance to complement-
 CC mediated opsonophagocytosis. The gene cluster is used as a source of
 CC probes and primers for serotype-specific detection of S. suis and is also
 CC useful for recombinant production of the proteins. The proteins are then
 CC useful for producing antigens that can be used in vaccines, for
 CC controlling or eradicating a Streptococcal disease, in humans or animals,
 CC e.g. against S. suis in pigs
 CC
 XX Sequence 338 AA;
 SQ

Query Match 57.2%; Score 1016.5; DB 3; Length 338;
 Best Local Similarity 59.2%; Pred. No. 2.2e-89;
 Matches 202; Conservative 45; Mismatches 89; Indels 5; Gaps 2;

QY 4 IYIVARICGNHNSVDIAEMILAKKAGVNAVFPQTKAKLISALPKAEYQIKN 63
 DB 2 IYIVARICGNHNSVDIAEMILAKKAGVNAVFPQTKAKLISALPKAEYQIKN 63
 QY 64 LBSQLEMTKKLEMKYDDYILHMEYAVSLNLDVFPSTPEDSIDFLASIKOKIMWIPS 123
 DB 62 LBSQLEMTKKLEMKYDDYILHMEYAVSLNLDVFPSTPEDSIDFLASIKOKIMWIPS 123

CC used in gene therapy. Antibodies to (I) are used for affinity chromatography, immunoassays, and distinguishing/identifying Streptococcus proteins

CC Sequence 341 AA:

Query Match 55.1%; Score 979.5; DB 5; Length 341;
Best Local Similarity 56.3%; Pred. No. 8.3e-86;
Matches 192; Conservative 54; Mismatches 90; Indels 5; Gaps 2;

CC 124 LNLPLEKIAKLPIPKKIIISTGMATIDBIKOSVIFINNKVPVGNITILHONTETPT 183
DB 122 TNLPLEKIGK--QKKVILSTGMVMEIHQAVNIIRKQGT--DISILHCTTEYPTP 176
QY 184 FEDVNLAINLNDKKEPPKNNIGPSDSSGFYAALAAVPGITFIEKFTLDSKMSGPDHL 243
DB 177 YPLNLNVNHTLKEEPNLTIGYSDSVSEVPILAAANGALEIEHFTLDMEMGPDHK 236
QY 244 ASIEPDELKHLCTGVRCVSKLSGNSKVTASERKKIYAKSIILAKTEIKGEVPSK 303
DB 237 ASATPDILALVKGRIVEQSLGKFEKEBEVEVERKKIYAKSIIVAKAIANGEVFTEN 296
QY 304 ITTKRPGNGISPMWYNIILGKIAEODPIFDELIIHSEFNQ 344
DB 297 ITVKRPGNGISPMWYNIILGQVSEQDFEDDNTCHSAFENQ 337

RESULT 5
ABP26810 standard; protein; 341 AA.

AC ABP26810:
DT 02-JUL-2002 (first entry)
XX Streptococcus polyepitide SEQ ID NO 2796.
XX Streptococcus polyepitide SEQ ID NO 2796.
XX Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
XX group A streptococcus; Streptococcus pyogenes; antibacterial;
XX anti-inflammatory; infection; vaccine; meningitis; gene therapy.
XX Streptococcus agalactiae.
XX MO200234771-A2.
XX 02-MAY-2002.
XX 29-OCT-2001; 2001WO-GB004789.
XX 27-OCT-2000; 2000GB-00026333.
XX 24-NOV-2000; 2000GB-00028727.
XX 07-MAR-2001; 2001GB-00005640.
XX (CHIR-) CHIRON SPA.
XX (GENO-) INST GENOMIC RES.
XX Telford J, Maignani V, Margalit Y Rost, Grandi G, Fraser C;
XX Tettelin H;
XX WPI, 2002-352536/38.
XX N-PSDB; ABN67441.
XX New Streptococcus protein for the treatment or prevention of infection or
XX disease caused by Streptococcus bacteria, such as meningitis, and for
XX detecting a compound that binds to the protein.
XX Claim 1; Page 3430; 4525pp; English.

CC The invention relates to a protein (ABP25413-ABP30895) from group B
CC Streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
CC the specification. The proteins have antibacterial and anti-inflammatory
CC activity. (I) nucleic acids encoding (II) ABN66044-ABN71526 and
CC antibodies that bind (I) are used in the manufacture of medicaments for
CC the treatment or prevention of infection or disease caused by
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC Nucleic acids encoding (I) are used to detect Streptococcus in a
CC biological sample. (I) is used to determine whether a compound binds to
CC (I). A composition comprising (I) or a nucleic acid encoding (II), may be
CC used as a vaccine or diagnostic composition. The disease caused by
CC Streptococcus that is prevented or treated may be meningitis. Nucleic
CC acid encoding (I) may be used to recombinantly produce (I) and may be

CC used in gene therapy. Antibodies to (I) are used for affinity chromatography, immunoassays, and distinguishing/identifying Streptococcus proteins

CC Sequence 341 AA:

Query Match 55.1%; Score 979.5; DB 5; Length 341;
Best Local Similarity 56.3%; Pred. No. 8.3e-86;
Matches 192; Conservative 54; Mismatches 90; Indels 5; Gaps 2;

CC 124 LNLPLEKIAKLPIPKKIIISTGMATIDBIKOSVIFINNKVPVGNITILHONTETPT 183
DB 122 TNLPLEKIGK--QKKVILSTGMVMEIHQAVNIIRKQGT--DISILHCTTEYPTP 176
QY 184 FEDVNLAINLNDKKEPPKNNIGPSDSSGFYAALAAVPGITFIEKFTLDSKMSGPDHL 243
DB 177 YPLNLNVNHTLKEEPNLTIGYSDSVSEVPILAAANGALEIEHFTLDMEMGPDHK 236
QY 244 ASIEPDELKHLCTGVRCVSKLSGNSKVTASERKKIYAKSIILAKTEIKGEVPSK 303
DB 237 ASATPDILALVKGRIVEQSLGKFEKEBEVEVERKKIYAKSIIVAKAIANGEVFTEN 296
QY 304 ITTKRPGNGISPMWYNIILGKIAEODPIFDELIIHSEFNQ 344
DB 297 ITVKRPGNGISPMWYNIILGQVSEQDFEDDNTCHSAFENQ 337

RESULT 6
ABP26810 standard; protein; 359 AA.

AC ABP26810:
DT 24-MAY-2002 (first entry)
XX Human polyepitide SEQ ID NO 2597.
XX Cytostatic; immunosuppressive; nocotropic; neuroprotective; antiviral;
XX anti-allergic; hepatotropic; antidiabetic; anti-inflammatory; antitumor;
XX vulnary; anticonvulsant; antibacterial; anti-fungal; antiparasitic;
XX cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
XX neurological disease; infection; human; secreted protein.
XX Homo sapiens.
XX WO200190304-A2.
XX 29-NOV-2001.
XX 18-MAY-2001; 2001WO-US016450.
XX 19-MAY-2000; 2000US-0205515P.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Birse CE, Rosen CA;
XX WPI, 2002-122018/16.
XX N-PSDB; ABP90630.
XX Novel 1405 isolated polypeptides, useful for diagnosis, treatment and
XX prevention of neural, immune system, muscular, reproductive,
XX gastrointestinal, pulmonary, cardiovascular, renal and proliferative
XX disorders.

PS Claim 11; SEQ ID NO 2597; 2081bp + Sequence Listing; English

The invention relates to novel genes (A3189449-43; B3190853) and proteins (A3B39040-A3B39044) useful for preventing, treating or ameliorating medical conditions e.g. by protein or gene therapy. The genes are isolated from a range of human tissues disclosed in the specification. The nucleic acids, proteins, antibodies and (anti)agonists are useful in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and ovarian cancer and other cancers of the adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune disorders e.g. Addison's disease, allergies, autoimmune haemolytic disorders; e.g. autoimmune thyroiditis, diabetes mellitus, Crohn's disease, anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c) cardiovascular disorders such as myocardial ischaemias; (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f) infectious diseases such as viral, bacterial, fungal and parasitic infections. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from http://atftp.wipo.int/pub/published_pct_sequences

Sequence 359 AA;

Query Match	29.0%	Score 516.5	DB 5	Length 359
Best Local Similarity	36.6%	Pred. No. 5.5e+41		
Matches 124; Conservative	61	Mismatches 139	Indels 15	Gaps 6

[illegible]

RESULT 7
AAM39986
ID AAM39986 standard; protein; 359 AA.

KM	Human; nocrotropic; immunosuppressant; cytostatic; gene therapy; cancer;
KM	peripheral nervous system; neuropathy; central nervous system; CNS;
KM	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KM	amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KM	chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KM	leukaemia.
XX	
XX	
OS	Homo sapiens.
XX	
PN	WO200153312-A1.
XX	
XX	
DD	26-JUL-2001.

XX. 26-DEC-2000; 2000WO-US034263.
PF

PR 23-DEC-1999, 99UTS-00471275,
PR 21-JAN-2000, 2000UTS-00468732,
PR 25-APR-2000, 2000UTS-00553371,
PR 20-JUN-2000, 2000UTS-00536042,
PR 19-JUL-2000, 2000UTS-00628013,
PR 03-AUG-2000, 2000UTS-00653445,
PR 14-SEP-2000, 2000UTS-00662191,
PR 19-OCT-2000, 2000UTS-00693036,
PR 29-NOV-2000, 2000UTS-00727344,

PA (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
PI Zhou P, Goodrich R, Drmanac RT;

DR WPI; 2001-442253/47.
DR N-PSDB; AA159142.

PT Novel nucleic acids and polypeptides, useful for treating disorders such as central nervous system injuries.

PS Example 4; SEQ ID NO 3131; 10078bp; English.

CC The invention relates to human nucleic acids (AA157798-AA161369) and the
CC encoded polypeptides (AA038642-AA042213) with neurotropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathies and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukemias and
CC C.N.S disorders. Note: The sequence data for this patent did not form
CC part of the printed specification

SQ Sequence 359 AA;

Query Match	28.9%	Score 514.5	DB 4	Length 359;
Best Local Similarity	36.6%	Pred. No. 8.6e-41;		
Matches 124; Conservative	60;	Mismatches 140;	Indels 15;	Gaps 6;

Qy	5	YVAAEIGCHNNSVLLAEMILKAEADVAAMVKQTEFADDLISALAPKAEYOQIKNTGEL	64	
		20	FIILAEIGNHODLVLAAKMLRMAEEGADCAKQTSLEKFKRKLDREPTYSKSHWG-	78
Dh				
Qy	65	ESOLENTKRLAKRYDDYLHMEYAVSLNLDVFSPPEDSDSDPLASIKOKIMKIPSGELL	124	
		79	KTYGEKRRLEPSHDQYRELQRYAAEYVGIPTLASGEMAEVFLHELNVPFFKVSQDYN	138
Dh				
Qy	125	NLPYLEKINAPLIPPKKLIISGCAATIDEIKOSSIPLNNKVPYGNITLTHCNEYEPTPF	184	
		139	NBPYLEKINAKKGRP---WVSSGQMSDMTMQVQYIV---KPLNPNCFLOCIISATYLOP	199
Dh				
Qy	185	EDVNINAINDLKAPPKKNIGSPDSHSGFYAALAAVPYGITPEKHEFTLDKMSGDPHLA	244	
		193	EDVNLRVISEYOKLPDPIDIGSGHGEGTIALSTAAVALAGKATVERHHTLLDKTWGSGHSA	252
Dh				
Qy	245	SEEPDELKHLICIGVACVEKSLGSSNSKYVTAASERKNKIVAKRSITAKRIKKGVSFSEKNI	304	
		253	SEEPDELAEVRSVTLVERALGSPFKOLLPEBMACNEKLSGVAKYKITEBGTILTMNML	312
Dh				
Qy	305	YTK--RPGNGISPMEMYNLLGK----IABODFPDELI	336	
		313	YTKVGEPR-KGYPPBDIFNLVGRKVLVTVVEEDDTIMEBIV	350
Dh				

RESULT 8
 AAB93183
 ID AAB93183 standard; protein, 359 AA.
 XX
 AC AAB93183;
 XX
 DT 26-JUN-2001 (first entry)
 XX
 DE Human protein sequence SEQ ID NO:12130.
 XX
 KM Human; primer; detection; diagnosis; antisense therapy; gene therapy.
 XX
 OS Homo sapiens.
 XX
 PN EPI074617-A2.
 XX
 PD 07-FEB-2001.
 XX
 PF 26-JUN-2000; 2000EP-00116126.
 XX
 PR 29-JUL-1999; 99JP-00248036.
 PR 27-AUG-1999; 99JP-0030253.
 PR 11-JAN-2000; 2000JP-00118776.
 PR 02-MAY-2000; 2000JP-00183767.
 PR 09-JUN-2000; 2000JP-00241899.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 DR WPI; 2001-318749/34.
 XX
 PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-
 PT length cDNAs defined in the specification, and for the detection and/or
 PT diagnosis of the abnormality of the proteins encoded by the full-length
 PT cDNAs.
 XX
 PS Claim 8; SEQ ID NO 12130; 2537pp + Sequence Listing; English.
 XX
 CC The present invention describes primer sets for synthesizing 5602 full-
 CC length cDNAs defined in the specification. Where a primer set comprises:
 CC (a) an oligo-dT primer and an oligonucleotide complementary to the
 CC complementary strand of a polynucleotide which comprises one of the 5602
 CC nucleotide sequences defined in the specification; where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence; where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in the
 CC specification. The primer sets can be used in antisense therapy and in
 CC gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any special methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
 CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
 CC oligonucleotides, all of which are used in the exemplification of the
 CC present invention
 XX
 SQ Sequence 359 AA;
 XX
 Query Match 28.9%; Score 514.5; DB 4; Length 359;
 Best Local Similarity 36.6%; Pred. No. 8.6e-41;
 Matches 124; Conservative 60; Mismatches 140; Indels 15; Gaps 6;
 QY 5 YVAAETGCGNNGSVDAAREMTAKKENGUNAVFQFFKADKLISALAPKAEYIKNTGEL 64
 DB 20 FIIAETGQNHGGLDVAKRMIRAKCGACCAKFKQSELEFFKRRKALDRPYTSKSSWG- 78

QY 65 ESQLEMTKTLKMKYKDYDLHMEYAVSLNDVSTPEDESDIFLASLKOKIKIPSGELL 124
 DB 79 KTYGEHRRHLEFSSHQYRELQRYAEFVCIFFLASGDEMAVFLHLELNPFPKVGSGDYN 138
 QY 125 NLPYLEKIKAKLPIPDKKIISTGMATIDEIKOSVSFFINNKKVPVGNITILHCNTEPTPE 184
 DB 139 NFPYLEKIKAKKGRP--WVTSQKQSDMTNKKVQYQV--KELNRPFCLOCTSAVFLQF 192
 QY 185 EDVNLNAINDLKHPKKNIGFSDHSSGFYAAIAAVPYGITEFKGFTLDKSMGPDHIA 244
 DB 193 EDVNLNVEISEYQKLFDPDIPIGSGHETGTAISVAVALGTKYLERHITLDKTKWGSDBSA 252
 QY 245 STEPDELKRLCIGYRVESKISGNSKVTASERKKIVAKKSLIKTEIKKEGVESEKNI 304
 DB 253 SLSEGLAELVRSVRLVERALGSPFKQLPCEMACNEKLGKSVAAKXLPBGITLMDWL 312
 QY 305 TTK--SPENGISPMEWYNLGR-----IAEODFIDELI 336
 DB 313 TVKVGSP-KGYPPEDIFNLVGGKVLVTFEEDDTIKRELV 350
 XX
 RESULT 9
 AAY96101
 ID AAY96101 standard; protein, 359 AA.
 XX
 AC AAY96101;
 XX
 DT 19-DEC-2000 (first entry)
 XX
 DE Human sialic acid synthetase.
 XX
 KM Sialic acid synthetase; human; sas gene; sialylation; glycoprotein;
 KM plasminogen; transferrin; thyrotropin; Na⁺/K⁺-ATPase.
 XX
 OS Homo sapiens.
 XX
 PN WO200052135-A2.
 XX
 PD 08-SEP-2000.
 XX
 PR 01-MAR-2000; 2000WO-US005313.
 XX
 PR 02-MAR-1999; 99US-0122582P.
 PR 08-DEC-1999; 99US-0169624P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 PA (UYUO-) UNIV JOHNS HOPKINS.
 PA (UYWY-) UNIV WYOMING.
 XX
 PI Betenbaugh MJ, Lawrence S, Lee YC, Jarvis D, Coleman TA;
 DR WPI; 2000-572178/53.
 DR N-PSDB; AAA50569.
 XX
 PT Recombinant production of sialylated glycoproteins using cells in which
 PT the expression of enzymes, e.g. sialic acid synthetase, involved in the
 PT sialylation reaction has been altered.
 XX
 XX Claim 16; Page 105-106; 144pp; English.
 XX
 CC The present sequence is that of human sialic acid synthetase (SAS), an
 CC enzyme that condenses Mann6-6-P or Man6-6-P with UDP to form Neu5Ac and
 CC KDN phosphates, respectively. The sequence was deduced from SAS cDNA (see
 CC AA50569). Northern blots indicated ubiquitous transcription of the SAS
 CC gene in a selection of tissues. The invention provides methods and
 CC recombinantly engineered cells for producing glycoproteins having
 CC sialylated oligosaccharides. The methods involve altering the expression
 CC of enzymes involved in carbohydrate processing. A claimed cell producing
 CC sialylated glycoprotein at above endogenous levels expresses at least 1
 CC (preferably human) enzyme selected from GlcNAc-2 epimerase, an enzyme
 CC catalyzing the conversion of UDP-GlcNAc to Mann6C, sialic acid
 CC synthetase, aldolase, CMP-sialic acid synthetase and CMP-sialic acid

[illegible]

Db	79	KTYGHEKRLLEESHOYELQRYAEVGIPTTASGDMVAEFLHMLNVEFFRVSGDPTN	138
Qy	125	NDEYLKIAKALDIPKRIIISTGATIDELKQSVISITNNKVPVGNITTHCHNTETPTP	164
Db	139	NPEYLKTKTKKRP---NWISGQMSMTKQVOIV---KPLNPNCFIQQCSAPPLOP	192
Qy	185	EDVNLAINIDLKQHPKONIGSPDSSGFVAALAAVPGITFPIEKHPTLDRKMSGPDHLA	244
Db	193	EDVNLAVISEYQKLFDDIPICSGHETIALSVAAALGAKVLERHITLIDKTKKSGDHA	252
Qy	245	SLEPDELKHLICIGRVCEKSLGSSNKVYTAASRRNKTIIVAKRSIIIAKTEIKKGEVPEKON	304
Db	253	SLEPGLAELAVSVLVERALGSPFKQLLPCEMAKNEKLGKSVAAKVIPEGITLITMDKL	312
Qy	305	TTK--RPGNGISPMENVNLCK-----IAEDDFIPELFI	336
Db	313	TVKVGEP-KAYPPEDFINLVGKNLVLYVEEDDTIMEELV	350
RESULT 12			
AAO26546			
AAO26546		standard, protein, 359 AA.	
XX	AAO26546,		
AC			
AD			
AE	06-MAR-2003 (first entry)		
AF			
AG			
AH	human SA synthetase protein.		
AI			
AK	Vaccine, glycoprotein, insect cell; enzyme: N-acetylglucosamine-2;		
AL	GlcNAc-2, epimerase; UDP-GlcNAc; mannose; (ManNAc; stalic acid;		
AM	synthetase; alolase; cytidine monophosphate-stalic acid; CMP-SA;		
AN	transporter; sialylated glycoprotein; human.		
AO			
AP	Homo sapiens.		
AQ			
AR	US2002142386-A1.		
AS			
AT	03-OCT-2002.		
AV			
AW	16-AUG-2001; 2001US-00930440.		
AX			
AY	02-MAR-1999; 99US-01225822.		
AZ	08-DEC-1999; 99US-0165624E.		
BA	25-AUG-2000; 2000US-0227579F.		
BB			
BC	(BETB/) BETENBAUGH M J.		
BD	(LAWR/) LAWRENCE S.		
BE	(LEEV/) LEE Y C.		
BF	(COLE/) COLEMAN T A.		
BG			
BH	Betenbaugh MJ, Lawrence S, Lee YC, Coleman TA;		
BI			
BJ	WPI; 2003-102519/09.		
BK			
BL	N-PSDB; AAL53993.		
BM			
BN			
BO			
BP	Manipulating glycoprotein production in insect cell, involves enhancing		
BQ	expression of enzymes involved in carbohydrate processing pathway such as		
BR	N-acetylglucosamine-2 epimerase or stalic acid synthetase.		
BS			
BT	Claim 16; Fig 32; 88pp; English.		
BU			
BV			
BW			
BX			
BY			
BZ			
CA			
CB			
CC			
CD			
CE			
CF			
CG			
CH			
CI			
CJ			
CK			
CL			
CM			
CN			
CO			
CP			
CQ			
CR			
CS			
CT			
CU			
CV			
CW			
CX			
CY			
CA			
CB			
CC			
CD			
CE			
CF			
CG			
CH			
CI			
CJ			
CK			
CL			
CM			
CN			
CO			
CP			
CQ			
CR			
CS			
CT			
CU			
CV			
CW			
CX			
CY			
CA			
CB			
CC			
CD			
CE			
CF			
CG			
CH			
CI			
CJ			
CK			
CL			
CM			
CN			
CO			
CP			
CQ			
CR			
CS			
CT			
CU			
CV			
CW			
CX			
CY			
CA			
CB			
CC			

CC useful as pharmaceutical compositions, vaccines, diagnostics and
 CC therapeutics. This sequence represents the human SA synthetase protein of
 CC the invention

XX Sequence 359 AA:

Query Match 28.7%; Score 510.5; DB 6; Length 359;
 Best Local Similarity 36.3%; Pred. No. 2.1e-40;
 Matches 123; Conservative 61; Mismatches 140; Indels 15; Gaps 6;

QY 5 YVAVICGNNGSVDIREMILKAKKGVNAVKQTRKADKLISALPKAEYQIKNGEL 64
 DB 20 FIIAEIQNNHGGDIDVAKRMIRAKKCCGACQAKFKSELEFKRKLERPYTSKSWG- 78
 QY 65 ESQLEMTKLEMKDYDLHMEYAVSINLDVSTPPEDESIDFLASIKOKIWKIPSGELL 124
 DB 79 KTVGEHGRHLEFSDQYRELQRYAEYGIFFTLASGDMVAFELHNLVPPFKVSGSDEN 138
 QY 125 NLPLEKIAKLPIDPKKIIISTGATIDEIKOSVITINNKVPVGNITIIHCNTEYTPF 184
 DB 139 NFPLEKTAKKGRP---WVISGQSMDFMKQVQIV---KPLNPNCFLOQTSAYPLQP 192
 QY 185 EDVNAINIDLKHPKNNIGFSDHSGFYAALAAVYGTIFIEKHFTLDKSMGPDHLA 244
 DB 193 EDVNAIRYISYQKLPDPIPIGSGHETGIALSVAAVLGAKVLERHITLDKTKGSDHSA 252
 QY 245 SIEPDELKHLCTIGRCVEKSLGNSKVVTAERKNTIARKSIIAKTEIKKGEVSEKNT 304
 DB 253 SLRGEIAELVRSVRLVERALGSPTRQLLPCENACNEKLGKSVAAKVIPEGITLLTMDKL 312
 QY 305 TTK--RPGNGSPHEWNLGK-----IAQDPIPDLLI 336
 DB 313 TVKVGEP-KAYPPEDITNLVGKVLTVVEBDTMEELV 350

RESULT 13

AA49716 ID AAM49716 standard; protein; 349 AA.

AC AAM49716;
 DT 07-JUN-2002 (first entry)
 XX

DE N. meningitidis sialic protein.

XX N-acetylneuraminic acid synthase; sialic acid; malignant cell;
 KW inflammatory process; cell surface glycoprotein; glycolipid;
 KM cell growth regulation; cell differentiation; metastasis.

XX Neisseria meningitidis.

XX DE10034586-A1.

XX 07-FEB-2002.

XX 14-JUL-2000; 2000DE-01034586.

XX 14-JUL-2000; 2000DE-01034586.

XX (FBSS/) FBSSNER W.

XX Fessner W, Krorst M;

XX WPI; 2002-29312/34.

XX N-PSDB; ABA99765.

XX High activity recombinant N-acetylneuraminic acid synthase production,
 PT for use in the preparation of N-acetylneuraminic acid compounds,
 PF including new derivatives and analogs.

XX Claim 4; Page 20-22; 26pp; German.

XX This invention describes a novel preparation of proteins having N-

CC acetylneuraminic acid synthase activity. The preparation involves
 CC culturing a prokaryotic host organism transformed by an expression vector
 CC containing a structural gene encoding proteins with N-acetylneuraminic
 CC acid synthase activity. The proteins of the invention are useful in the
 CC production of sialic acids and their derivatives and analogues. Sialic
 CC acids are components of cell surface glycoproteins and glycolipids, and
 CC are involved in physiological and pathological recognition processes,
 CC e.g. in inflammatory processes, regulation of cell growth and
 CC differentiation, malignant cell dedifferentiation and metastasis or as
 CC recognition domains of pathological viruses and bacteria. Glycoproteins
 CC and glycolipids based on non-natural sialic acid derivatives and
 CC analogues may be useful as metabolically stable therapeutic agents or as
 CC fluorescently labeled analytical/diagnostic agents. The proteins obtained
 CC by the present method are more accessible than prior art analogues, and
 CC have significantly higher specific activity, markedly wider substrate
 CC tolerance and superior stability. This sequence represents the Neisseria
 CC meningitidis sialic gene described in the disclosure of the invention

XX Sequence 349 AA:

Query Match 28.0%; Score 497.5; DB 5; Length 349;
 Best Local Similarity 37.3%; Pred. No. 3.6e-39;
 Matches 120; Conservative 50; Mismatches 141; Indels 11; Gaps 5;

QY 6 TVAEICGNNGSVDIREMILKAKKGVNAVKQTRKADKLISALPKAEYQIKNGELE 65
 DB 22 IIEICINHEGSLKTFEWDAAVNAKAEVVKQTHIVE---DEMDKQVIRGADV- 77
 QY 66 SQLEMTKLEMKDYDLHMEYAVSINLDVSTPPEDESIDFLASIKOKIWKIPSGELLN 125
 DB 78 SIYEIMERKALNEDEIKLKEAYESKMTFISTPFRALRLQMDIPAYKIGSECCNN 137
 QY 126 LPLEKIAKLPIDPKKIIISTGATIDEIKOSVITINNKVPVGNITIIHCNTEYTPPE 185
 DB 138 YPLIKLVASF--GKPIILSTGNSTIESIKSVETIIRKGV--YALHCTNYITPTIS 191
 QY 186 DVNAINIDLKHPKNNIGFSDHSGFYAALAAVYGTIFIEKHFTLDKSMGPDHLAS 245
 DB 192 DVRLGMDLSEAFPAIIGLSDHLDNVAACGAVALGSIIEHRTDNRDRGPVICS 251
 QY 246 IEPDELKHLCTIGRCVEKSLGNSKVVTAERKNTIARKSIIAKTEIKKGEVSEKNT 305
 DB 252 MNPDTFKELKQGMALAKARGKQDTIIGEKPTDPAPASVADDDIKKGLSDGNLM 311
 QY 306 TKRPGNG-ISPHEWNLGKIA 326
 DB 312 VKRPGNGDPSVNEVETLPGRVA 333

RESULT 14

AA41772 ID AAM41772 standard; protein; 400 AA.

AC AAM41772;

DT 22-OCT-2001 (first entry)

XX Human polypeptide SEQ ID NO 6703.

XX Human; noctropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; hemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemocastic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia.

XX Homo sapiens.

XX WO200153312-A1.

XX 26-JUL-2001.

XX 26-DEC-2000; 2000WO-US034263.

Thu May 13 11:53:13 2004

us-09-930-440b-8.rag

Page 11

QY 0 IYAEIGCHNSGVDAIAREHILKAKENAGVNAKFOFPADKILSLAPKAYOIKNTGELE 65
 Db 19 VPEIIGINSGSEIMKILWDAWASTGAKITIGQTHIWEDESSGAQKV---IPNAKI- 74
 QY 66 SOLEMTKLEMKYDDYLHLMEYASVNLDFSTPEDESDIDFLASLKOKIMKIPSEGLLN 125
 Db 75 SIYEIMQKCALYKOLALKEYTEKIGVYISTEPISAGARLEDMQVSAFKIGSECCNN 134
 QY 126 LPFLKIALKLPIDPKKIIISTGMATIDELKOSVSIFINKKPVGNITILACHTIEPTPEE 185
 Db 135 YPLKIKIAA---KKEMIVSTGMSISEIKPTVILIDNEIP---PYLMETMLYPTPPN 188
 QY 186 DVNLNAINDLKCHFFPKCNKNGFSDDHSGGYAALAAVPGVITIEHFTLDSXSGDPDHLAS 245
 Db 189 LVNLNMLLEIKKEF--SCWGLSDHDTTDNLACIGAAALGACVLEHFPDMSMRGPIVOS 247
 QY 246 IEEDDELKHCIGVRCVENLSGNS--SKVVTASERKKIYARKSIIIAKTEIKKGVSEFENKI 304
 Db 248 MDTQALKEILIOEQALMRGNNSKKAQKQOYVITIDFAFASVYSIDIKKGVLSMDNI 307
 QY 305 TTKRPG--NGISPMWENYILGKLAEDQIPIDELIHSFE 341
 Db 308 WYKRPGLGGISIAAEFENILGKQALNDIENDTOLSTEDB 345

Search completed: May 6, 2004, 09:04:27
Job time : 52.7787 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:00:50 ; Search time 13.1424 Seconds
(without alignments)
2532.427 Million cell updates/sec

Title: US-09-930-440B-8

Sequence: 1 MSNIIYVAEIGCNHNSVDI.....EQDIPDELIHSFKNQGE 346

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	1778	100.0	346	2	169836
2	979.5	55.1	341	2	T44651
3	867	48.8	334	2	D81276
4	647	36.4	337	2	H64432
5	504	28.3	359	2	IC7321
6	497.5	28.0	349	2	SC6760
7	493	27.7	350	2	F97169
8	492.5	27.7	343	2	B81275
9	472	26.5	373	2	S39722
10	443.5	24.9	356	2	D87604
11	426	24.0	343	2	H81318
12	402	22.6	340	2	B54542
13	381	21.4	340	2	B71965
14	266.5	15.0	360	2	H71307
15	124	7.0	678	2	E90595
16	113	6.4	1222	2	B90593
17	110.5	6.2	1365	1	BVBYS
18	110	6.2	769	2	SS4525
19	109.5	6.2	420	2	B84788
20	109	6.1	601	2	C81715
21	107.5	6.0	713	2	S64437
22	107.5	6.0	2748	2	S57976
23	107	6.0	409	2	T43599
24	106.5	6.0	1577	2	T19722
25	106.5	6.0	1061	2	C86630
26	105.5	5.9	1426	2	SC2439
27	105.5	5.9	1620	2	S61535
28	103.5	5.8	623	2	S56206
29	103	5.8	653	2	A29821

30	103	5.8	732	2	UC6099	Ku autoantigen 86k
31	103	5.8	950	2	E70203	exonuclease Sbc {
32	103	5.8	2269	2	T28677	thoxytry protein -
33	102.5	5.8	519	2	E82932	spermidine/putresc
34	102.5	5.8	567	2	F64453	oxalacetate decar
35	102.5	5.8	739	2	JN0581	vascular cell adhe
36	102.5	5.8	1163	2	D64315	type I restriction
37	102.5	5.8	1186	2	C64588	cag pathogenicity
38	102	5.7	1036	2	E96882	hypothetical prote
39	102	5.7	1241	2	F97286	DNA-dependent RNA
40	102	5.7	2136	2	A05037	hypothetical prote
41	101	5.7	857	2	E95009	cell wall surface
42	101	5.7	4688	2	F82885	hypothetical prote
43	100.5	5.7	342	2	H81317	probable lipopolys
44	100.5	5.7	563	2	B82883	hypothetical prote
45	100	5.6	557	2	B83962	hypothetical prote

ALIGNMENTS

RESULT 1

169836
neub protein - Escherichia coli

C/Species: Escherichia coli

C/Date: 07-Jun-1996 #sequence_reviseion 07-Jun-1996 #text_change 20-Jun-2000

C/Accession: 169836

R:Annunziato, P.W.; Wright, L.F.; Vann, W.F.; Silver, R.P.

J. Bacteriol. 177, 312-319, 1995

A/Title: Nucleotide sequence and genetic analysis of the nub and nub genes in region

A/Reference number: 155145; MIMD:5513767; PMID:7814319

A/Accession: 169836

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1346 <RES>

A/Cross-References: EMBL:U05248; NID:9454079; PID:NAC43302.1; PID:9454081

C/Superfamily: Streptococcus agalactiae capsular polysaccharide biosynthesis protein cpi

Query Match 100.0%; Score 1778; DB 2; Length 346;
Best Local Similarity 100.0%; Pred. No. 1.7e-117;
Matches 346; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MSNIIYVAEIGCNHNSVDIARENIIKAKAGVNAVKFCFFKADKLISAIAPAEYQIKN	60
DB	1	MSNIIYVAEIGCNHNSVDIARENIIKAKAGVNAVKFCFFKADKLISAIAPAEYQIKN	60
QY	61	TGEIESOLEMTKKLEMYDYDYLHMEYAVSLNDFVSTPPDESDIDFLASKOKIKWIPS	120
DB	61	TGEIESOLEMTKKLEMYDYDYLHMEYAVSLNDFVSTPPDESDIDFLASKOKIKWIPS	120
QY	121	GELLNLPYLKIAKLPIDPKIIISTGMATIDIKOSVSIFINNKPVGNIITLHCITEY	180
DB	121	GELLNLPYLKIAKLPIDPKIIISTGMATIDIKOSVSIFINNKPVGNIITLHCITEY	180
QY	181	PTPEFDVNAINDLKKHPKXNIGFSDHSGGYAIAAVPYGITFEKFTLDKMSGP	240
DB	181	PTPEFDVNAINDLKKHPKXNIGFSDHSGGYAIAAVPYGITFEKFTLDKMSGP	240
QY	241	DHLASIEPDELKILCIQVRCVSKSNGSVYASERKXKIVAKSIIARTETIKGCVFS	300
DB	241	DHLASIEPDELKILCIQVRCVSKSNGSVYASERKXKIVAKSIIARTETIKGCVFS	300
QY	301	EKKITTRPGNGISPMEWTLLGKIAQDPIPELIHSFKNQGE	346
DB	301	EKKITTRPGNGISPMEWTLLGKIAQDPIPELIHSFKNQGE	346

RESULT 2

T44651
capsular polysaccharide biosynthesis protein cpsN (imported) - Streptococcus agalactiae

C/Species: Streptococcus agalactiae

C/Date: 21-Jan-2000 #sequence_reviseion 21-Jan-2000 #text_change 09-Jun-2000

C/Accession: T44651

R:Chaffin, D.O.; Yim, H.H.; Beres, S.B.; Sweet, E.S.; Nitayajarn, A.; Rubens, C.E.
Submitted to the EMBL Data Library, June 1999

A:Accession: T44651
A:Reference number: 222821
A:Status: preliminary; translated from GB/EMBL/DBD
A:Molecule type: DNA
A:Residues: 1-341 <CHA>
A:Cross-references: EMBL:AF163833; PIDN:AAB53074.1
A:Experimental source: strain COH1, serotype III
C:Genetics:
A:Gene: cpsN
C:Superfamily: Streptococcus agalactiae capsular polysaccharide biosynthesis protein cps

Query Match 55.1%; Score 979.5; DB 2; Length 341;
Best Local Similarity 56.3%; Pred. No. 26-61;
Matches 192; Conservative 54; Mismatches 90; Indels 5; Gaps 2;

4 IYVAFGCHNGSVDAIEEMILKAEAGNAVKPOTFKADKLISALAPKAYQINNGE 63
2 VYVIAEGCHNGSDINAKKAVVAWSCGVDAVKPOTFKAEKLISFAPAPAEYQKATYGT 61
64 LESOLEMTKLEMKYDDYLTLMAYAVSLNDVSTPDESDIDFLASLKOKIWKIPSGEL 123
62 AI QLEMTKLEHSEFYLEMNDYALSKGVETSTPDESDLEFLISTOMPIKIPSGEI 121
DB 124 LKPYLEKAKLPIPKKIITSTGATIDIKOSVSIPIFNKVPVGNITTLCHNTYPTP 183
122 TMLPYLEKIGK--QCKVILSTGMVMEIHQAVNLKONGTT--DISILCTTEYPTP 176
184 FEVNNALINDLKCHFPKXNNIGSDHSGFYAALAVPYGITPIEKGFTLDKSMGPDHL 243
177 YPSLNLNVHTLTKDEFKDLITGSDHSISEVPILAAAGCAVIEHGFILDTMBGPDHK 236
DB 244 ASIEPELKLHLCTGVPCVEKSGDSNRKVTASERKNTIYARSKIATKEIKGEVPSKN 303
237 ASATPDLIALVGVAVIVEQALCFEPIKIPVPEKRIYARSKVVALKIKKQDIVSIEN 296
QY 304 ITTKRPGNGISPMWYNLGLKIAEODFIPELTIHSEPPNQ 344
DB 297 ITVKRPGNGISPMWYNLGLKIAEODFIPELTIHSEPPNQ 337

RESULT 3
DB1276
N-acetylneuraminic acid synthetase (EC 4.1.3.-) C1327 (imported) - Campylobacter jejuni

C:Species: Campylobacter jejuni
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 03-Jun-2002
C:Accession: DB1276
R:Parkhill, J.; Wren, B.W.; Mangall, K.; Kelsey, G.M.; Churcher, C.; Basham, D.; Chilling
C.W.; Quail, M.; Randleam, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Barrell
Nature 403, 665-668, 2000
A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hnt
A:Reference number: AB1250; NMID:20150912; PMID:10688204
A:Accession: DB1276
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-334 <PAR>
A:Cross-references: GB:AL139078; GB:AL111168; NID:G6968723; PIDN:CAB73754.1; PID:G696878
A:Experimental source: serotype O2, strain NCTC 1168
C:Genetics:
A:Gene: neuB2; C1327
C:Superfamily: Streptococcus agalactiae capsular polysaccharide biosynthesis protein cps
C:Keywords: Carbon-carbon lyase; oxo-acid-lyase

Query Match 48.8%; Score 867; DB 2; Length 334;
Best Local Similarity 51.5%; Pred. No. 1.5e-53;
Matches 173; Conservative 62; Mismatches 97; Indels 4; Gaps 2;
QY 1 MSNIVVAEIGCHNGSVDAIEEMILKAEAGNAVKPOTFKADKLISALAPKAYQIN 60
DB 1 MKKTLIAEAGVNHNDINLAKLIEIAADSGADPVKFSFKAKNCISTKAKKAPYQTKT 60
QY 61 TGELESQLKEMTKLEMKYDDYLTLMAYAVSLNDVSTPDESDIDFLASLKOKIWKIPS 120

DB 61 TANDESQGWQCKELDKELDKAEKELIAKKNIAFLSTPDLSEVDLNEGLKFKIDS 120

QY 121 GELLNLPYLEKAKLPIPKKIITSTGATIDIKOSVSIPIFNKVPVGNITTLCHNTY 180
DB 121 GELLNLPYLEKAKLPIPKKIITSTGATIDIKOSVSIPIFNKVPVGNITTLCHNTY 177
QY 181 PTFEDVALINDLKCHFPKXNNIGSDHSGFYAALAVPYGITPIEKGFTLDKSMGSP 240
DB 178 PAFNENVLKAMQSLKDAF-KLDVGYSDHTGIRHISLAVALGACVIEKHFTLDKSMGSP 236

QY 241 DMLAIEZDELKHLCTGVPCVEKSGDSNRKVTASERKNTIYARSKIATKEIKGEVPS 300
DB 237 DPKALEZDELKHLCTGVPCVEKSGDSNRKVTASERKNTIYARSKIATKEIKGEVPS 296

QY 301 EKNITTKRPGNGISPMWYNLGLKIAEODFIPELTIHSEPPNQ 346
DB 297 EGNITTKRPGNGISPMWYNLGLKIAEODFIPELTIHSEPPNQ 332

RESULT 4
HE6432
spore coat polysaccharide biosynthesis protein E homolog - Methanococcus jannaschii

C:Species: Methanococcus jannaschii
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
C:Accession: HE6432
R:Bull, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,
; Reisch, C.J.; Overbeek, R.; Kirkness, E.B.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.
; Ison, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hirst, M.A.
Science 273, 1058-1073, 1996
A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Moese,
A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii
A:Reference number: A64300; NMID:96337999; PMID:8688087
A:Accession: HE6432
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-337 <BDU>
A:Cross-references: GB:U67549; GB:U77117; NID:G2826363; PIDN:AAB59068.1; PID:G1591717;
A:Map position: REV1006919-1005906
C:Superfamily: Streptococcus agalactiae capsular polysaccharide biosynthesis protein cps

Query Match 36.4%; Score 647; DB 2; Length 337;
Best Local Similarity 42.2%; Pred. No. 4.4e-38;
Matches 140; Conservative 56; Mismatches 116; Indels 20; Gaps 5;

QY 5 YVIAEGCHNGSVDAIEEMILKAEAGNAVKPOTFKADKLISALAPKAYQINTEL 64
DB 18 FLIAGSINHGNDIDIGELVNAKCAQADAKRQSYHTBDFIS--KKSEY----- 66
QY 65 ESQLEMTKLEMKYDDYLTLMAYAVSLNDVSTPDESDIDFLASLKOKIWKIPSGEL 124
DB 67 --YELFSLSESESEFELKAVKIGIMFISFPLDKYVDILNRKAVPAFKIASGDLT 123
QY 125 NLPLYEKAKLPIPKKIITSTGATIDIKOSVSIPIFNKVPVGNITTLCHNTYPTP 184
DB 124 FYPLLEKAKLPIPKKIITSTGATIDIKOSVSIPIFNKVPVGNITTLCHNTYPTP 178
QY 185 EDVNNALINDLKCHFPKXNNIGSDHSGFYAALAVPYGITPIEKGFTLDKSMGPDHL 244
DB 179 EDVNNALINDLKCHFPKXNNIGSDHSGFYAALAVPYGITPIEKGFTLDKSMGPDHL 237
QY 245 STPEPELHLCTGVPCVEKSGDSNRKVTASERKNTIYARSKIATKEIKGEVPSKN 304
DB 238 SADPEPELHLCTGVPCVEKSGDSNRKVTASERKNTIYARSKIATKEIKGEVPSKN 297
QY 305 TTKRPGNGISPMWYNLGLKIAEODFIPELTIHSEPPNQ 346
DB 298 SFKRGNGISPMWYNLGLKIAEODFIPELTIHSEPPNQ 332

RESULT 5
JC7321

N-acetylneuraminic acid 9-phosphate synthase (EC 4.1.3.-) - mouse
 NAlternate names: N-acetylneuraminic acid synthase homolog
 C:Species: Mus musculus (house mouse)
 C:Date: 08-Sep-2000 #sequence_revision 08-Sep-2000 #text_change 06-Oct-2000
 C:Accession: J07321
 R:Nakata, D.; Close, B.E.; Colley, K.J.; Matsuda, T.; Kitaajima, K.
 Biochem. Biophys. Res. Commun. 273, 642-648, 2000
 A:Title: Molecular cloning and expression of the mouse N-acetylneuraminic acid 9-phosphate
 A:Reference number: J07321
 A:Accession: J07321
 A:Residues: 1-359 <NAK>
 A:Molecule type: mRNA
 A:Cross-references: DDBJ:AB041263
 A:Comment: This enzyme is a cytosolic enzyme and as a housekeeping enzyme as well, which
 catalyzes the conversion of N-acetylneuraminic acid 9-phosphate
 C:Genetics:
 A:Gene: static acid
 C:Superfamily: Streptococcus agalactiae capsular polysaccharide biosynthesis protein cps
 C:Keywords: carbon-carbon lyase; oxo-acid-lyase

Query Match 28.3%; Score 504; DB 2; Length 359;
 Best Local Similarity 36.0%; Pred. No. 5.3e-28;
 Matches 124; Conservative 59; Mismatches 145; Indels 16; Gaps 7;

QY 5 YVVAIGCNHNSVDIAEMTLKAKAGVNAVKEQTFKADKLISALAPKAEYQIKNTGEL 64
 DB 20 FLIAEIGNHGDIDVAKMIRTAECGACDCAKQKSELEKFNKALEREPTSKHSG- 78
 QY 65 ESQLEMTKLEMKYDDYVLMMEYASVNLDPFSTPDEDSIDFLASLKQIKTKIPSGEL 124
 DB 79 KYGKHKLHLSHSHQYELQVLAQELGIFFTAGMDMAVEFHELVNPPFKAGSGGTN 138
 QY 125 NLPLYEKIAKLPIPKKIIISTGATIDELIKOSVIFNNKVPVGNITILHCNTEYPTPE 184
 DB 139 NPLYEKIAKAKGRP---WISSGQSMQMTKQVQIV---KLMPNPFQCTATVLP 192
 QY 185 EDVNLAINDLKQHPKKNIGSPDHSGCFYALAAVPGITFIEKHFTLDKMSGPDHIA 244
 DB 193 EDANLRVISEYQKFPDPPIGYSGHETGIALSVAAVLAQAVLRHTLDTWYSGDHS 252
 QY 245 SIEPDELKHLICGRCEKSGKSNKYVTASERKNKYIARSKITAEIKKGEVSEKNT 304
 DB 253 SIEPDELAEVSRVLRALASFPKQDLPCEMANKGSGSVARAKKIPAGTTLTIDTL 312
 QY 305 TTK-RPGNGISPMEMYLLAK---IAQDFIPDELI-THSE 340
 DB 313 TVKVGEP-KGYPPEDIFYLAKKVLVTEEDDTWMESEVESHK 355

RESULT 6

S60760
 polygalactonic acid capsule biosynthesis protein Slc NMB0068 [imported] - Neisseria meningi-
 C:Species: Neisseria meningitidis
 C:Date: 27-Apr-1996 #sequence_revision 13-Mar-1997 #text_change 19-Jan-2001
 C:Accession: S60760; B81241
 R:Edwards, U.; Mueller, A.; Hammer Schmidt, S.; Gerardy-Schahn, R.; Froesch, M.
 Mol. Microbiol. 14, 141-149, 1994
 A:Title: Molecular analysis of the biosynthesis pathway of the alpha-2,8 polysialic acid
 A:Reference number: S60760; PMID:95131727; PMID:7830552
 A:Accession: S60760
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-349 <EDW>
 A:Cross-references: EMBL:M59053; NID:9520732; PIDN:AAA20477.1; PID:9520735
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1994
 R:Teitelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.B.; Eisen, J.A.
 Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;
 Ri, H.; Olin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Massigani, V.; Pizsa, M.
 A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappelli, R.; Ve
 A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain Mc58.
 A:Reference number: A51000; MUID:20175755; PMID:10710307
 A:Accession: B81241

A:Molecule type: DNA
 A:Residues: 1349 <IFT>
 A:Cross-references: GB:AE002366; GB:AE002098; NID:97225284; PIDN:AAF40535.1; PID:972252
 A:Experimental source: serogroup B, strain Mc58
 C:Genetics:
 A:Gene: NMB0068
 C:Superfamily: Streptococcus agalactiae capsular polysaccharide biosynthesis protein cp

Query Match 28.0%; Score 497.5; DB 2; Length 349;
 Best Local Similarity 37.3%; Pred. No. 1.5e-27;
 Matches 120; Conservative 50; Mismatches 141; Indels 11; Gaps 5;

QY 6 IVAIGCNHNSVDIAEMTLKAKAGVNAVKEQTFKADKLISALAPKAEYQIKNTGEL 65
 DB 22 IICGIGNHGSILKTAEDVAVNAGAEVYKQTHVE---DEMSDEAKQVIRNADV- 77
 QY 66 ESQLEMTKLEMKYDDYVLMMEYASVNLDPFSTPDEDSIDFLASLKQIKTKIPSGEL 125
 DB 78 SIVEIMERKALNEDBEIKLAEVSEKMFISTPFSRAALRLQRMIDIPAYKIGSGCNN 137
 QY 126 LPLYEKIAKLPIPKKIIISTGATIDELIKOSVIFNNKVPVGNITILHCNTEYPTPE 185
 DB 138 YPLIKLVASF---GKPIILSTGANSIESIKSYEILREAGVP---YALHCTNLYPTPE 191
 QY 186 DVNLAINDLKQHPKKNIGSPDHSGCFYALAAVPGITFIEKHFTLDKMSGPDHIA 245
 DB 192 DVLRGMNDLSEAFPDALIGLSDHTLDNLYACLAVALGSGILSRHFTDRMRDREPDIWCS 251
 QY 246 IEPDELKHLICGRCEKSGKSNKYVTASERKNKYIARSKITAEIKKGEVSEKNT 305
 DB 252 KNPPTFELKQGAHALALAGKQKOTIIAEKPTKQAFASVADKDKGSLSGNLM 311
 QY 306 TKRPGNG-ISPMEWYLLAKIA 326
 DB 312 VKRQNGDPSVNEVELFGKVA 333

RESULT 7

P97169
 static acid synthase [imported] - Clostridium acetobutylicum
 C:Species: Clostridium acetobutylicum
 C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 30-Sep-2001
 C:Accession: P97169
 R:Nolling, U.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
 J. Bacteriol. 183, 4823-4838, 2001
 A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Cl.
 A:Reference number: A96900; MUID:21359325; PMID:21359325
 A:Accession: P97169
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-350 <KUR>
 A:Cross-references: GB:AE001437; PIDN:AAK80145.1; PID:G15025183; GSPDB:GN00168
 A:Experimental source: Clostridium acetobutylicum ATCC824
 C:Genetics:
 A:Gene: CAC2187
 C:Superfamily: Streptococcus agalactiae capsular polysaccharide biosynthesis protein cps

Query Match 27.7%; Score 493; DB 2; Length 350;

Best Local Similarity 35.7%; Pred. No. 3e-27;
 Matches 117; Conservative 69; Mismatches 130; Indels 12; Gaps 7;

QY 5 YVVAIGCNHNSVDIAEMTLKAKAGVNAVKEQTFKADKLISALAPKAEYQIKNTGEL 64
 DB 19 FLIAEIGNHGDIDVAKMIRTAECGACDCAKQKSELEKFNKALEREPTSKHSG-CTI 76
 QY 65 ESQLEMTKLEMKYDDYVLMMEYASVNLDPFSTPDEDSIDFLASLKQIKTKIPSGEL 121
 DB 77 MDGTTLHKLYEEAVTQWQPKLEIAEEGGLCFSPDPNTSVDFLERNEVAYKVASF 136
 QY 122 ELNLNPLYEKIAKLPIPKKIIISTGATIDELIKOSVIFNNKVPVGNITILHCNTEYPT 181
 DB 137 ELNDIPPIEIIAS---KGRPVINSTGIALENGSTIYEAVAAC--RQGENVITLLKCSSTP 191

QY 182 TPEEDVNLAINDKKHPKNNIGFSDHSGFYAAIAVPYGI:FTPEKHTLTKSMSPD 241
 192 SPEDINLKTIPMKREAF-NCSVGLSDHTMGYSVANAALATVIEKFTLKRSDDGPD 250
 QY 242 HIASIPDELKHLCTGRCVCKSLGNSKRVTAERKNIVAKSIIATETIKGSEFSE 301
 251 SAFSMPPEESAAVAKSIREVEKALGKVTYELTEKQNSRSHSLFVVR-DIKKGETFTK 309
 QY 302 KNITTKRPGKISPPMEMYMLGKIAEOD 329
 310 ENKVSIRPAFGLTKYIEEVIGKKAARD 337
 Db

RESULT 8
 B81275
 N-acetylneuraminic acid synthetase (EC 4.1.3.-) Cj1317 (imported) - Campylobacter jejuni
 C/Species: Campylobacter jejuni
 C/Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #ext_change 03-Jun-2002
 C/Accession: B81275
 R:Parkhill, J.; Wren, B.W.; Mungall, K.; Kelsey, J.M.; Churcher, C.; Basham, D.; Chilling
 C.W.; Quail, M.; Randleam, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Barrall
 Nature 403, 665-668, 2000
 A>Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyf
 A/Reference number: A81250; MUID:20150912; PMID:10688204
 A/Accession: B81275
 A/Status: Preliminary
 A/Molecule type: DNA
 A/Residues: 1-343 <PAR>
 A/Cross-references: GB:AL113078; GB:AL111168; NID:G6968723; PINN:CA873744.1; PID:G696875
 A/Experimental source: serotype O2, strain NCTC 11168
 C/Genetics:
 A/Gene: neuB3, Cj1317
 C/Superfamily: Streptococcus agalactiae capsular polysaccharide biosynthesis protein cps
 C/Keywords: carbon-carbon lyase; oxo-acid-lyase

Query Match 27.7%; Score 492.5; DB 2; Length 343;
 Best Local Similarity 35.3%; Pred. No. 3.2e-27;
 Matches 120; Conservative 67; Mismatches 144; Indels 9; Gaps 6;

QY 4 IYVAETGCGNNGSVDIAREMILKAEAGVNAVRFQTFKADKLISALAPKAEYQIK-NTG 62
 12 VFIIAEISANHAGSLTMAKLSIKAKAGADAIKIQTYPDST-TINSDEKEDPIIKGGLM 70
 Db
 QY 63 ELSSQLEMTKLEMKYDDVLIHMEVAVSLNLDVFSFPEDESDIDFLASKQIKIMKIPGE 122
 71 DKRKLYELVESATPYEMHSQIFETLONIGILCFSSPKMEDEVEFLKRPDPAYKIASFE 130
 Db
 QY 123 LNLPLYLEKIAKLPIPDKKIISTGMATIDELIKOSYSIFENNVKVPNGNTIILHCNTEYPT 182
 131 ANDENFVRLIAK--EKKPTIVSTGATTEELFKICEIFEBEKNP--DVLFKCTSTYPT 185
 Db
 QY 183 PFEDVNLAINDKKHPKNNIGFSDHSGFYAAIAVPYGI:FTPEKHTLTKSMSPD 242
 186 AIEDMMKLGIVSKEKF-NVEVGLSDHSGFYAAIAVPYGI:FTPEKHTLTKSMSPD 244
 Db
 QY 243 LASIEPDELKHLCTGRCVCKSLGNSKRVTAERKNIVAKSIIATETIKGSEFSEK 302
 245 KPSLDIDPEFPAWDAVRAQBSALGDKDLDEKVLKNRYPAR-SLYASNDIKKSEMFSE 303
 Db
 QY 303 NITTKRPGKISPPMEMYMLGKIAEODPIPEDELIIHSEK 342
 304 NVKVSIRPAFGLTKYIEEVIGKKAARD 343
 Db

RESULT 9
 S39722
 spore coat polysaccharide synthetis protein spse - Bacillus subtilis
 N/Alternate names: protein ipa-660
 C/Species: Bacillus subtilis
 C/Date: 07-Oct-1994 #sequence_revision 26-May-1995 #ext_change 20-Jun-2000
 C/Accession: S39722; C69717
 R/Glaser, P.; Kunst, F.; Arnaut, M.; Coudart, M.P.; Gonzales, W.; Hullo, M.F.; Ionescu,

A.; Rapoport, G.; Danchin, A.
 Mol. Microbiol. 10, 371-384, 1993
 A>Title: Bacillus subtilis genome project: cloning and sequencing of the 97 kb region f
 A/Reference number: S39655; MUID:95020537; PMID:7534828
 A/Accession: S39722
 A/Status: Preliminary; nucleic acid sequence not shown; translation not shown
 A/Molecule type: DNA
 A/Residues: 1-373 <GLA>

A/Cross-references: EMBL:X73124; NID:q413923; PINN:CA451623.1; PID:q413991
 A/Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1993
 R/Kunst, F.; Ogatawara, M.; Moszer, I.; Albertin, A.M.; Alloni, G.; Azevedo, V.; Berte
 C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capano, V.; Carter, N.M.; Ch
 A.; Enrllich, S.D.; Emerson, P.T.; Enlhan, K.D.; Errington, J.; Fabret, C.; Ferrati, E.
 Nature 390, 249-256, 1997
 A/Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Funai, S.; Galizzi, A.; Galie
 jech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holtsappel, S.; Hosono, S.; Hullo, M.
 Koetter, J.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidos, A.; Lardinois
 A/Authors: Lamber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue
 Y., M.; Ogawa, K.; Ogatawara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Potetel
 Rieger, W.; Rivolta, A.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon
 A/Authors: Schleich, S.; Schroeder, R.; Scottone, F.; Sekiguchi, J.; Sekowska, A.; Senc
 akuch, M.; Tamakoshi, A.; Tanaka, T.; Terpetra, P.; Tognoni, A.; Tosato, V.; Uchiyama
 T.; Winters, P.; Wipet, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida,
 A/Authors: Yoshikawa, H.F.; Zumestein, B.; Yoshikawa, H.; Danchin, A.
 A>Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
 A/Reference number: A69580; MUID:98044033; PMID:9384377
 A/Accession: C69717
 A/Status: Preliminary; nucleic acid sequence not shown; translation not shown
 A/Molecule type: DNA
 A/Residues: 1-373 <KUN>
 A/Cross-references: GB:Z99123; GB:AL009126; NID:G2636240; PIDN:CA815813.1; PID:G2636324
 A/Experimental source: strain 168
 C/Genetics:
 A/Gene: spsB
 C/Superfamily: Streptococcus agalactiae capsular polysaccharide biosynthesis protein ci

Query Match 26.5%; Score 472; DB 2; Length 373;
 Best Local Similarity 36.4%; Pred. No. 9.8e-26;
 Matches 132; Conservative 55; Mismatches 128; Indels 48; Gaps 11;

QY 4 IYVAETGCGNNGSVDIAREMILKAEAGVNAVRFQTFKADKLISALAPKAEYQIKNTGE 63
 17 VFIIAEISANHAGSLTMAKLSIKAKAGADAIKIQTYPDST-TINSDEKEDPIIKGGLM 70
 Db
 QY 64 LE-----SOLEMTKLEMKYDDVLIHMEVAVSLNLDVFSFPEDESDIDFLASKQIKIM 116
 67 YKTAAGKQVSTFSLVQSKEMPAEWLPIPLDTCREKQVIFLSTVCDEGADLQSTSPAF 126
 Db
 QY 117 KIPSGELNPLYLEKIAKLPIPDKKIISTGMATIDELIKO--SVSIFINKVPNGNTI 173
 127 KIASYEINHLPLKIVARL--NRPMISTGAEISDVHEAMRTIRAGNNO-----TAI 178
 Db
 QY 174 LHCNTEYTPPEEDVNLAINDKKHPKNNIGFSDHSGFYAAIAVPYGI:FTPEKHT 232
 179 MECAVAPAPPEYSLSVTPYLAAPFAVAVIGFSDHSHPLPAPCAARLQKLIKIF 238
 Db
 QY 233 LDKSMGPDHSLASIEPDELKHLCTGVR-----CVKSLGNSKRVTAERKN 279
 239 IDKULPGADHSPALNPDCLKMVGIRKTBALQKIGITKAPSEKLLGSSYKTTVAIMEEI 298
 Db
 QY 280 KIVAKSIIATETIKGSEFSEKNTTTRPG--NGISPPMEMYML--GKIAEODPIPE 334
 299 RNFAPRGIFTTAPQIKGSAFSEDMITAVLRPQKQGLHP-RFELLTSGVRAVRIDPDT 357
 Db

RESULT 10
 D87604
 new protein, probable [imported] - Caulobacter crescentus
 C/Species: Caulobacter crescentus

C>Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
 C/Accession: D87604
 R:Manan, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.B.; Leub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolton, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M. Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
 A>Title: Complete Genome Sequence of *Caulobacter crescentus*.
 A/Reference number: A87249; MUID:21173698; PMID:11259647
 A/Accession: D87604
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-356 <STO>
 A/Cross-references: GB:AE005673; NID:913424484; PIDN:AAK24832.1; GSPDB:GN00148
 C/Genetics:
 A:Gene: CC2868
 C:Superfamily: Streptococcus agalactiae capsular polysaccharide biosynthesis protein cps

Query Match 24.9%; Score 443.5; DB 2; Length 356;
 Best Local Similarity 32.5%; Pred. No. 9.2e-24;
 Matches 109; Conservative 68; Mismatches 141; Indels 17; Gaps 8;
 QY 5 YIVAEICNNGSVDIAREMILKAEAGVAVNKKPQTEKADKLISAIAPKAEYQIKNTGEL 64
 26 YVIEELSGNNGSIEKCIAMVDAADTCDAIKIQTITAD-TITLDVDRFEKIH--GSL 82
 QY 65 --ESQLEMTKLEMKYDYILHMEYAVSLNLDVSTFPDESIDFLASKOKIKWIPSG 121
 DB 83 WDGETLVELYEAEATPEPWEHAIEPBARORGATFFSSPDETAADLDSIGAEFFKIASF 142
 QY 122 ELMLPHEIKAKLPIDDKKIISTGMATIDEMQSSIFINMKVPKGNITLHCNTEYP 181
 DB 143 EAVDLPPIKIAA--AKGKFLISTGMALTEKWTALDPAISGAP--GVLLHCVSSTP 197
 QY 182 TPEEDVNLAINDIKKKPF--PKNNIGSDSSGFFAAIAAVPYGITEKHFITLDSKMSG 239
 DB 198 ATFADAVVRIVPDAARFGCP---IGLSDHTPGTAASVAAVSLGACAVEKHFLLRADGG 254
 QY 240 PDHLASIEPEDELKELICGVRCVSKSGNSKVTYASRKKIKVARKSIITIKIKGEVF 299
 DB 255 PDAFSLPEAPFKALVDOTKAAALAPAHYDVAGSS-APSLFRSLVTADVKAGEPL 313
 QY 300 SEKNITTPKPGNGISPMEWYNLLGKIAEODFIPDE 334
 DB 314 TRAVRSVRPGNGLPADLDKVLGKATRLDARGE 348

RESULT 11

H81318
 N-acetylneuraminic acid synthetase (EC 4.1.3.-) Cj1141 [imported] - *Campylobacter jejuni*
 C/Species: *Campylobacter jejuni*
 C/Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 03-Jun-2002
 C/Accession: H81318
 R:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chilling, C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Barrett, Nature 403, 665-668, 2000
 A>Title: The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hyf
 A/Reference number: A81250; MUID:20150912; PMID:10688204
 A/Accession: H81318
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-343 <PAR>
 A/Cross-references: GB:AL139077; GB:AL111168; NID:86968444; PIDN:CA873396.1; PID:9696857
 A:Experimental source: serotype O2, strain NCTC 11168
 C/Genetics:
 A:Gene: Cj1141
 C:Superfamily: Streptococcus agalactiae capsular polysaccharide biosynthesis protein cps
 C/Keywords: carbon-carbon lyase; oxo-acid-lyase

Query Match 24.0%; Score 426; DB 2; Length 343;
 Best Local Similarity 35.0%; Pred. No. 1.5e-22;
 Matches 115; Conservative 57; Mismatches 143; Indels 14; Gaps 7;

3 NIVVAIEGCHNNGSVDIAREMILKAEAGVAVNKKPQTEKADKLISAIAPKAEYQIKNTG 62

DB 14 NPIILPEIGINNGSLKIAKLVADAKRAGAKIKNTGTHIVE---DEMGEAKVPIPGNA 70
 QY 63 ELESQLEMTKLEMKYDYILHMEYAVSLNLDVSTFPDESIDFLASKOKIKWIPSGE 122
 DB 71 NI-SIEIMEQALNKFDLAKAEYKQGLVYLSFERRAANRLEDGVSAYKGSGB 129
 QY 123 LNLPLYEIKAKLPIDDKKIISTGMATIDEMQSSIFINMKVPKGNITLHCNTEYP 182
 DB 130 CNNYPLIKIAQ---KKMIISTGMNLSIEIKPTKILRDYIEF---FVLLHTTLLYPT 183
 QY 183 PREDVNLAINDIKKKFPKNNI-GPSDSSGFFAAIAAVPYGITEKHFITLDSKMSGD 241
 DB 184 PSHLVRLQAMLELYKEF--NCLYGLSDHTTNLACIGALAGASVLERHFTDMRKRPD 241
 QY 242 HIASIEPEDELKELICGVRCVSKSGNSKVTYASRKKIKVARKSIITIKIKGEVFSE 301
 DB 242 IVCSNDSSTLKDILNQTQEMVLKSGNNKPKLEBYITDPAFASVSIKDIKGEILSM 301
 QY 302 KNITTRPG-NGISPMEWYNLLGKIAEOD 329
 DB 302 DNIMVVRPSKGGISANDFEALIGKRAKD 330

RESULT 12

B64542
 spore coat polysaccharide biosynthesis protein B - *Helicobacter pylori* (strain 26695)
 C/Species: *Helicobacter pylori*
 C/Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 20-Jun-2000
 C/Accession: B64542
 R:Tombs, J.F.; White, O.; Kervatage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.; Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khajik, H.G.; Glodek, A.; McKern, son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watney, L. Nature 388, 539-547, 1997
 A:Authors: Wallin, B.; Hayes, W.S.; Borodovsky, M.; Karp, P.D.; Smith, H.O.; Fraser, C. A>Title: The complete genome sequence of the gastric pathogen *Helicobacter pylori*.
 A/Reference number: A64520; MUID:97394467; PMID:9252185
 A/Accession: B64542
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown
 A/Molecule type: DNA
 A/Residues: 1-340 <TOM>
 A/Cross-references: GB:AE005328; GB:AE005111; NID:92313263; PIDN:AAD07248.1; PID:9231326
 C:Superfamily: Streptococcus agalactiae capsular polysaccharide biosynthesis protein cps

Query Match 22.6%; Score 402; DB 2; Length 340;
 Best Local Similarity 31.8%; Pred. No. 7.1e-21;
 Matches 104; Conservative 74; Mismatches 135; Indels 14; Gaps 8;

QY 6 IVAEICNNGSVDIAREMILKAEAGVAVNKKPQTEKADKLISAIAPKAEYQIKNTG 64
 DB 7 IVAEILANNHODLNAKESLHAKESGADPVKIQIT-TSCMTLNKEDOPFIQITLMOK 65
 QY 65 ESQLEMTKLEMKYDYILHMEYAVSLNLDVSTFPDESIDFLASKOKIKWIPSGEEL 124
 DB 66 ENLYEYORASTPLEVHAELFEIARLDDIGTESPSSQALBELLSINCVMYKIASFEIV 125
 QY 125 NLYPEKIAKLPIDDKKIISTGMATIDEMQSSIF--INNKVPKGNITLHCNTEYP 182
 DB 126 DDLILKAAK---TQKPIILSSGIAHTHLODASISCRANN---FDITLAKVAYS 178
 QY 183 PREDVNLAINDIKKKFPKNNIGFSDSSGFFAAIAAVPYGITEKHFITLDSKMSGPD 242
 DB 179 KIDANMLLSVKYKGIYGV-KFGLSDHTIGSLCPILATITGASMIKHPITLNSLOTPDS 237
 QY 243 LASIEPEDELKELICGVRCVSKSGNS-KVYASRKKIKVARKSIITIKIKGEVFSE 301
 DB 238 AFSVDNFGKSWEAIKQSVLALGSEEPRIINPKTLAKRPFASLVYIK-DIQKGEALTR 296
 QY 302 KNITTRPGNGISPMEWYNLLGKIAEOD 328
 DB 297 NNKALRPMLGLHPKRYEITIGQKAK 323

RESULT 13

B71965
static acid synthase - Helicobacter pylori (strain J99)

C:Species: Helicobacter pylori

A:Variety: strain J99

C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 20-Jun-2000

C:Accession: B71965

R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, S.D.; Doig, P.C.; Smith, D.R.;

Ives, C.; Gibson, R.; Werberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;

Nature 397, 176-180, 1999

A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric path

A:Reference number: A71800; MID:99120557; PMID:9923682

A:Accession: B71965

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-340 <ARN>

A:Cross-references: GB:AE001455; GB:AE001439; NID:G4154678; PIDN:AA05747.1; PID:G415468

A:Experimental source: strain J99

C:Genetics:

A:Gene: neuB

C:Superfamily: Streptococcus agalactiae capsular polysaccharide biosynthesis protein cps

Query Match 21.4%; Score 381; DB 2; Length 340;

Best Local Similarity 30.1%; Pred. No. 2.1e-19;

Matches 98; Conservative 75; Mismatches 141; Indels 12; Gaps 6;

6 IVAEICGNHNSVDIAREMILKAKAGVNAVRFOTFRADKLISAIAPAEYOIKNT-GSL 64

7 IVAEISNHNQDINLAESLHAKESGADFKQIT-TRSCMLDSKEDPFIIQGTLMXK 65

65 ESQLETKLEMKYDDYTHMEYAVSLNLFSTPDEDSIDFLASKIKIMKIPSGEL 124

66 ENYGYQKASTPLEHNAHLELAKKLDIGIFSSPSKALLESIDCPMYKIASFEIV 125

125 NPEYLEKIAKLPIDPKKIIISTGMATIDELKOSVSIF--INNRVPVGNITIIHNTERT 182

126 DLEIKKAAK--TQRPILISSGIAIHAELQDASLCRGVNN---FDITLKCVAAYS 178

183 PFEDVNAIINLIDKKEFPKNNIGFSDHSGFYAIAAVPGITFIEKHFTLDKMSGDPH 242

179 KIDPAHLISNVLGATPGV-KFGSDHTIGSLCPILATLIGASMIKHLNKS-LTPDS 237

243 LASIEPELKHLCIGVRCVEKISGNSKVTATSERKKIVAKRSIIAKTEIKGEVFSK 302

238 AFEMDNGFMSVGAIKQSVLAGESEPKINPTLIERRFPALFVIDIQGELTSD 297

303 NITTKRPGNGISPMEMVNLKIAEQ 328

298 NIKALRPNIQALHFKFYKILQKASK 323

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Query Match 15.0%; Score 266.5; DB 2; Length 378;

Best Local Similarity 26.1%; Pred. No. 2.7e-11;

Matches 90; Conservative 60; Mismatches 152; Indels 43; Gaps 8;

2 SNIVYVAGCNHNSVDIAREMILKAKAGVNAVRFOTFRADKLISAIAPAEYOIKNT 61

13 ADULTIAGSHAGSPRABALIDAADAAAANKPFLIYAHLEHLPL-----T 62

62 GELE-----SOLETKLEMKYDDYTHMEYAVSLNLFSTPDEDSIDFLASKIKI 115

63 GAVRLPSGAVSLYQRFEELEVPISFYAACFPHARSGRLVIGISPPGPSATEALAKDF 122

116 KRIPEGELNIPYEKIAKLPIDPKKIIISTGMATIDELKOSVSIFINNRVPVGNITIIH 175

123 LKVASPE-LNYFTL-ISTLAALPILSSGVCILKIBGALACRCRYTQGGSHALH 179

176 CHTETPPEFEDVNAIINDLKKHFPKNNIGFSDHSGFYAIAAVPGITFIEKHFTLD 234

180 CITAYPAPBTEYNALALPALATIF-NINWGVSDHSDVDLVPFLARAGACIVEKHICLS 238

235 KMSGPDHLASIEPELKHLCIGVRCVEKISGNSKVTATSERKKIVAKRSIIAKTEIK 272

239 RTDAGLSDISALDPPDFRTPALNSCARREPSQIISFLHKGVAIPHVRAVIGSGEVL 298

273 TASEKNNIVAKRSIIAKTEIKGEVFSKNTTKRPGNGISPM 317

299 APERAAHYQKSNRSLHYLAHVPRTVLQKENVLYRSEANLSAGE 343

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Thu May 13 11:53:14 2004

us-09-930-440b-8.rpr

Page 7

Cy 269 AKTEIKKGEVSEKNITTKRPGNGISPMENYULGKIAEODPIPELTHSEF 341
D5 530 QISGFRITIDEFIRONFSSL--GRAIRTRKEGFDLKNYSA-SDLIPREDIPASAY 579

Search completed: May 6, 2004, 09:09:48
Job time : 16.3924 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 08:56:04 ; Search time 9.09861 Seconds
(without alignments)
1980.112 Million cell updates/sec

Title: US-09-930-440b-8

Perfect score: 1778
Sequence: 1 MSNIVVAEIGCNHNSVDI.....KQDFIDELIHSEFNQGE 346

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	647	36.4	337	1	YAE5_METUA
2	516.5	29.0	359	1	SIAS_HUMAN
3	472	26.5	373	1	SPSE_BACSU
4	110.5	6.2	1365	1	KRES_YEAST
5	110	6.2	769	1	MLAH_YEAST
6	109	6.1	601	1	ISPG_CHAMU
7	107.5	6.0	713	1	YGS8_YEAST
8	107.5	6.0	2748	1	NUM1_YEAST
9	107	6.0	409	1	YOPM_YEPE
10	105.5	5.9	1428	1	YAB4_SCHPO
11	103.5	5.8	623	1	YEP3_YEAST
12	102.5	5.8	567	1	PYC8_YEAST
13	102.5	5.8	739	1	YCAL_MOUSE
14	102.5	5.8	1075	1	Y124_METUA
15	102.5	5.8	1186	1	CAGA_HELPY
16	102	5.7	1241	1	RPOB_CLOAB
17	102	5.7	2136	1	YCF2_MARPO
18	100	5.6	878	1	MSH4_YEAST
19	99.5	5.6	887	1	TOP1_BACAN
20	98.5	5.5	373	1	ASPM_MOUSE
21	98.5	5.5	654	1	PGKT_THEMA
22	97.5	5.5	332	1	NDPA_VIBPA
23	97.5	5.5	962	1	IFJ3A_MAIZE
24	97.5	5.5	969	1	DP3A_URSPA
25	97.5	5.5	1460	1	N159_YEAST
26	97.5	5.5	5171	1	BPBA_HUMAN
27	97	5.5	554	1	Y478_RICPR
28	96.5	5.4	409	1	Y4DM_RHISN
29	96.5	5.4	1138	1	CTAA_BACTU
30	96.5	5.4	1782	1	VIT_BOMMO
31	96.5	5.4	1862	1	CSAB_PICPA
32	96	5.4	426	1	PUR2_FUSNM
33	96	5.4	603	1	UVRC_LISMO

34	96	5.4	1169	1	SMC_METUA	059037 methanococ
35	95	5.3	375	1	FLG1_BUCBP	089485 buchnera ap
36	94.5	5.3	597	1	NFL2_RAT	054968 rattus norv
37	94.5	5.3	646	1	SR72_YEAST	P38688 saccharomyc
38	94.5	5.3	2667	1	RBP2_PLAVB	000799 plasmodium
39	94	5.3	336	1	Y625_METUA	058042 methanococ
40	94	5.3	395	1	PURK_STNP7	054978 synecococc
41	94	5.3	501	1	MTBB_BACSU	P33563 bacillus su
42	94	5.3	798	1	SYEP_BUCAP	P59057 buchnera ap
43	94	5.3	1583	1	GCC2_HUMAN	081WJ2 homo sapien
44	93.5	5.3	415	1	Y310_BUCAP	08K9M4 buchnera ap
45	93.5	5.3	729	1	KAR3_YEAST	P17119 saccharomyc

ALIGNMENTS

RESULT 1
ID YAE5_METUA STANDARD; PRT; 337 AA.
AC Q58465;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein MJ1065.
GN MJ1065.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxId=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688067;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Nguyen D.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhman J.D., Sadow P.W., Hanna M.C.,
RA Uitterlinden L.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
jannaschii.";
RL Science 273:1058-1073(1996).
CC - SIMILARITY: STRONG, TO B. SUBTILIS SPSE.
CC - SIMILARITY: Contains 1 AFP-like domain.
CC - SIMILARITY: Contains 1 AFP-like domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
DR EMBL: U67549; AAC99068.1; -
DR PIR: H64432; H64432.
DR TRIR: MJ1065; -
DR InterPro: IPR006014; Antifreeze dom.
DR InterPro: IPR006190; Antifreeze-like.
DR InterPro: IPR004144; Neut.
DR Pfam: PF01354; Antifreeze; 1.
DR Pfam: PF03102; Neut; 1.
DR PROSITE: PS50844; AFP-LIKE; 1.
DR Hypothetical protein; Complete proteome.
FT DOMAIN 279 337 AFP-LIKE.
SQ SEQUENCE 337 AA; 37976 MW; 790CCAF4C111B CRC64;

Query Match 36.4%; Score 647; DB 1; Length 337;
Best local similarity 42.2%; Pred. No. 3.7e-37;
Matches 140; Conservative 56; Mismatches 116; Indels 20; Gaps 5;

RA Faley J, Helton E., Kettelman A.C., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Bakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywnski M.I., Skelske U., Smallus D.E.,
RA Schermer A., Schein J.E., Jones S.J., Marra M.A.,
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences." ,
RL Proc. Natl. Acad. Sci. U.S.A. 99:16699-16903(2002).
CC -1- FUNCTION: Produces N-acetylneuraminic acid (Neu5Ac) and 2-keto-3-
CC deoxy-D-lyxero-D-galacto-nononic acid (KDN). Can also use N-
CC acetylmannosamine 6-phosphate and mannoside 6-phosphate as
CC substrates to generate phosphorylated forms of Neu5Ac and KDN,
CC respectively.
CC -1- CATALYTIC ACTIVITY: Phosphoenolpyruvate + N-acetyl-D-mannosamine +
CC H(2)O = phosphate + N-acetylneuraminic acid.
CC -1- CATALYTIC ACTIVITY: Phosphoenolpyruvate + N-acetyl-D-mannosamine 6-
CC phosphate + H(2)O = N-acetylneuraminic acid + phosphate.
CC -1- TISSUE SPECIFICITY: Ubiquitous.
CC -1- SIMILARITY: Contains 1 Apr-like domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation at
CC the Bioinformatics Resource Centre. There are no restrictions on its

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement>
CC or send an email to license@sib-sib.ch).

DR	EMBL; AK001659; BAA8183.1; -
DR	EMBL; BC000008; AA000008.1; -
DR	EMBL; BC019315; AAH19315.1; -
DR	MIM; 605202; -
DR	HSSD; P19614; 9AME.
DR	GO; GO:0005737; C:Cytoplasm; NAS.
DR	GO; GO:0008783; F:N-acetylneuraminate cytidyltransferase acti. -; NAS.
DR	GO; GO:0009103; F:flippopolyaccharide biosynthesis; NAS.
DR	InterPro; IPR006104; Antifreeze_dom.
DR	InterPro; IPR006190; Antifreeze_like.
DR	InterPro; IPR004144; Neub.
DR	Pfam; PF03154; Antifreeze; 1.
DR	Pfam; PF03102; Neub; 1.
DR	ProDom; PD003258; Antifreezeit1; 1.
DR	PROSITE; PS00844; AFP_LIKE; 1.
IW	Transferase; Polymorphism.
FT	DOMAIN 294 353
FT	VARIANT 68 68
FT	E-> D /AFP-LIKE.
FT	CONFLICT 232 232 /FTID-VAR_01308.
FT	A-> T (IN REF. 2).

```

SQ      SEQUENCE 359 AA, 40307 MW, 26025 Da (RefSeq: NC007777)
Query Match      29.0%; Score 516.5; DB 1; Length 359;
Best Local Similarity 36.6%; Pred. No. 3.2e-28;
Matches 124; Conservative 61; Mismatches 139; Indels 15; Gaps 6;

QY      YVAAIGGNNHGSVDIAREMTLKAKBAEAVNVVKTQTFADKLISAIAPKAELQIKNTL 64
DB      20 FIATIGGNNHGDVDVAKMTIRAKKACADCAKQKSLERKFKRKALERTYTSHGSG- 78

```

Qy 65 ESQJMTKREMKRQDDYHMLHRYAVALDWFSIFEDUSJDFEASLQKQTMALF8GJAL 124

Db 79 KTYGHEKHLEFFSHQYRELQRYAEVGIFFETASGMDMAVEFLHMLVPPFXVSGGTN 138

Qy 125 NLPYLEKTAJLPIPDKXIIISTGMAITDRIKOSVIFLNKVPQGNITILCNTEYPPPF 184

Db 139 NPPYLEKTAJGRP--WVSSGQSMKTMKOVQIV--KPLPNPFCLQCTSIYPPQOP 192

Qy 185 BDVNLNAINDLKKHFPKNNKIGFSDHSSGFAVAIAAPVYGITFIEKHFTLDKSMSPDILA 244

Db 193 BDVNLRVATSEYQKLPFDIPIGYSGHETGATLSVAAPVALGAKVLEKHITLTDKTMKSDHSA 252

CC 245 SIEPELHLCIGVRCVCKSGNSKVYASERKNKLVARKSIITAKTEIKGEVSEKNI 304
 CC 253 SLEEGELAEIVRSVRLVRLALGSLFKQLPCPMACNEKXGSKSVAKVPIEGTILTLMDML 312
 CC 305 TTK--RQNGSGSPMWNVTLGK-----IAEQDFIDEI 336
 CC 313 TVKVGEP-KGYPPEDIFNLVCKKLVLYVEEDDTIMEELV 350
 CC Db

RESULT 3
 SPSE_BACSU STANDARD; PRT: 373 AA.
 AC P39625.1
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Spore coat polysaccharide biosynthesis protein spse.
 GN SPSE OR IPA-67D OR BSU37870.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxId=1423;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=95020537; PubMed=7934828;
 RA Glaser P., Kunst F., Arnaut M., Coudart M.P., Gonzales W.,
 RA Hulio M.F., Ionescu M., Ludochinsky B., Marcelino L., Moszer I.,
 RA Presseau E., Santana M., Schneider E., Schweizer J., Vertes A.,
 RA Rapoport G., Danchin A.;
 RT "Bacillus subtilis genome project: cloning and sequencing of the 97
 RT kb region from 325 degrees to 333 degrees";
 RL Mol. Microbiol. 10:371-384(1993).
 RL (2)
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=95044033; PubMed=9384377;
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
 RA Azevedo V., Bartorelli L., Bessieres P., Bolotin A., Borchert S.,
 RA Borries R., Boursier L., Brans A., Braun M., Brigelli S.C., Bron S.,
 RA Brouillet S., Brunsch C.V., Caldwell B., Capuano Y., Carter N.M.,
 RA Choi S.K., Codan J.J., Conterton I.F., Cummings N.J., Daniel R.A.,
 RA Danczoz F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
 RA Enright K.D., Erington J., Fabret C., Ferrari E., Foulger D.,
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
 RA Ghuu S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
 RA Guisepi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
 RA Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,
 RA Joris B., Karamata D., Kaashara Y., Klaerr-Blanchard M., Klein C.,
 RA Kobayashi Y., Koester P., Koningsstein G., Krogh S., Kumano M.,
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
 RA Pairo V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.K.,
 RA Presseau E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
 RA Rieger M., Rivoita C., Rocha B., Roche M., Rose M., Sadie Y.,
 RA Sato F., Scanlan E., Schleich S., Schroeter R., Scofield F.,
 RA Setiguchi J., Sekoweki A., Seror S.J., Serro P., Shin B.S., Soldo B.,
 RA Sorokin A., Taconi E., Takagi T., Takahashi H., Takemaru K.,
 RA Takeuchi M., Tanakoshi A., Tanaka T., Terpetra P., Tognoni A.,
 RA Totsato V., Uchiyama S., Vandenbol M., Vannier F., Vassatotti A.,
 RA Varti A., Wandut R., Wedler B., Wedler H., Weltzenegger T.,
 RA Winters P., Wipit A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zimstein E., Yoshikawa H., Danchin A.;
 RT "The complete genome sequence of the Gram-positive bacterium Bacillus
 RT subtilis";
 RL Nature 390:249-256(1997).
 RL -1- PATHWAY: Spore coat polysaccharide biosynthesis.
 CC -1- SIMILARITY: Contains 1 AFP-like domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X73124; CA51623.1; -
 CC EMBL: Z99123; CB1813.1; -
 CC PIR: S39722; S39722.
 CC Subtilin: BGI0613; spse.
 CC InterPro: IPR006014; Antifreeze_dom.
 CC InterPro: IPR006190; Antifreeze_like.
 CC InterPro: IPR004144; Neut.
 CC Pfam: PF01354; Antifreeze_1.
 CC Pfam: PF01102; Neut.
 CC PROSITE: PS50844; AFP_LIKE; 1.
 CC Complete proteome.
 CC DOMAIN 305 367 AFP-LIKE.
 CC SEQUENCE 373 AA; 40889 MW; 3CFBDFACB0DBCE8 CRC64;
 CC SO

Query Match 26.5%; Score 472; DB 1; Length 373;
 Best Local Similarity 36.4%; Pred. No. 3.6e-25;
 Matches 132; Conservative 55; Mismatches 128; Indels 48; Gaps 11;

CC 4 IYVAEIGCHNGSVDIAREMILKAKAGVAVAFOTFRADKLISALAPYAEYQIKNTGE 63
 CC 17 VFLIAEAGINHDKLQAFALIDAAEAGADVAFQMFQADRM-----YQ-KDPL 66
 CC 64 LE-----SQLEMTKLEKTDYHLMEYAVSLNDVSTFPDESIDPLASLKRW 116
 CC 67 YKTAGKGVSVFISVQSMEMPAEWIPLADYCKEKOVIPLSTVCDBSADLSTSPSAF 126
 CC 117 KIPSGELNLPYBIKIAKIPDKIISTGMATIDEIKO---SVSIFNNKYPVGNTI 173
 CC 127 KINSYEINHLPLKYARL---NRPMFSTAGAFISDVHEAMETIAEGNQ-----IAI 178
 CC 174 LACNTEYPTPEFDVNINAINDLKHPKKNIGFSDHS--GFVAIAAVPYGITEHNT 232
 CC 179 MGVAKPAPPEYSNLSVIFMLAAAPPEAVTGFSDSHPTPEAPCAVIRGAKLIEHFT 238
 CC 233 LDKSMGPDHLAIEPDELHLCIGVR-----CVKSGNSKVYASERKN 279
 CC 239 IDNLPADSPALNDELKEMVDGIRKTEALKOGITPVSKLSSYKTTTALBGI 286
 CC 280 KIVAKSIITAKTEIKGEVSEKNIITKRG--NGISPMWYNTL--GKIAEQDFIDE 334
 CC 299 RNFAVYGIPTTAIOGGAFAFSEENIAVLRCQKPGGLHP-RFPELLTSGVAVADIPADT 357
 CC 335 LIT 337
 CC 358 GIV 360
 CC Db

RESULT 4
 KRES_YEAST STANDARD; PRT: 1365 AA.
 ID KRES_YEAST
 AC P22023; Q12190;
 DT 01-AUG-1991 (Rel. 19, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Killer toxin-resistance protein 5 precursor.
 DE Killer toxin-resistance protein 5 precursor.
 GN KRES OR YOR336W.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Saccharomycetes; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 CC Eukaryota; Fungi; Ascomycota; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxId=4932;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90258892; PubMed=2188106;
 RA Meaden P., Hill K., Wagner J., Slipetz D., Sommer S.S., Bussey H.;
 RT "The yeast KRES gene encodes a probable endoplasmic reticulum protein
 RT required for (1->6)-beta-D-glucan synthesis and normal cell growth";
 RL Mol. Cell. Biol. 10:3013-3019(1990).
 RL (2)

RP SEQUENCE FROM N.A.
 RX MEDLINE=97051586; PubMed=8896263;
 RA Parle-Mcdermott A.G., Hand N.J., Goulding S.G., Wolfe K.H.;
 RT "Sequence of 29 kb around the PPR10 locus on the right arm of
 RT Saccharomyces cerevisiae chromosome XV: similarity to part of
 RT chromosome I.";
 RL Yeast 12:999-1004(1996).
 CC -1- FUNCTION: Required for (1->6)-beta-D-glucan synthesis and normal
 CC cell growth.
 CC -1- SUBCELLULAR LOCATION: Endoplasmic reticulum lumen.
 CC -1- SIMILARITY: SOME, TO D.MELANOGASTER USGG.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M33556; AAA34725.1; -;
 DR EMBL; 249821; CAA89981.1; -;
 DR EMBL; 275244; CAA99659.1; -;
 DR PIR; S62066; BVBHYK. -;
 DR GenBank; 143924; -;
 DR SCD; S0005863; KRES. -;
 DR GO; GO:0003980; F:UDP-glucose:glycoprotein glucosyltransferase. .; IDA.
 DR InterPro; IPR000886; ER target S.
 DR PROSITE; PS00014; ER TARGET; 1.
 DR KW Cell wall; Endoplasmic reticulum; signal; glycoprotein.
 FT SIGNAL 1 17
 FT CHAIN 1 17
 FT CARBOHYD 115 1365
 FT CARBOHYD 115 115
 FT CARBOHYD 228 228
 FT CARBOHYD 293 293
 FT CARBOHYD 457 457
 FT CARBOHYD 519 519
 FT CARBOHYD 523 523
 FT CARBOHYD 644 644
 FT CARBOHYD 870 870
 FT CARBOHYD 1091 1091
 FT CARBOHYD 1150 1150
 FT CARBOHYD 1195 1195
 FT SITE 1362 1365
 FT SITE 582 582
 FT CONFLICT 780 794
 FT SEQUENCE 1365 AA; 156476 MW; D0F5851175C033 CRC64;
 Query Match 6.2%; Score 110.5; DB 1; Length 1365;
 Best Local Similarity 21.5%; Pred. No. 7.6;
 Matches 78; Conservative 60; Mismatches 127; Indels 97; Gaps 19;
 QY 62 GELBSQLEMTK--KLEKDYDYLMHAYANSLNLFVSTPPEDESDIDFLASLQKWKIP 119
 DB 41 GADVNDQ-LTNLYPLVTGDEDEIDQENVALNSVLRERYDEVDADLELVAASY--P 97
 QY 120 SGEI-----LNTPELEKIAKLPIPD-----KKITITSGMATIDE 153
 DB 98 MGNMIGHNDSNAEDDANNSSYPVINGNRK-----PDVFLFKSDLDLTQKQVDV 151
 QY 154 IK-QSVSIFLNKKVPGNITILHCNTEYPPPEVDNIN---AIN-DLKHFPKNNIGS 207
 DB 152 IQPYDVAVIGNNSKAPI--LILYGCPTVIDSDPEFEFNFLFMEAMNGKGRFTWRSTCS 209
 QY 208 DHSSEGF-----ATAAVPYGITFEK-HFTLDKSM---SGPDHLASIEPD 249
 DB 210 DGRSVEYPLTHPLTITLQNGSRKSIPO---LKKILYTPKELVAGANDDQHLHLEP 265
 QY 250 ELAHLICGVACVEKSLASNSKVYASERKKIVAR-----KSIANTBIKKEVSEKN 303
 DB 266 ELRELDLRVSLISEFYQYKKDITATILNFTKSIIVNPPILSKDILIKSVNKKIITSNEE 325

QY 304 ITTK-----RPGNSISPMEH-----YNLIG-----KIAE--QDPIDELLHS 339
 DB 326 LMSKGFDMGLGLYINGQNMKTSILTPNLLTALNTEYQSLKIKTNLLQELPEPSCIIDS 385
 QY 340 EF 341
 DB 366 KF 387
 RESULT 5
 ID MLH1_YEAST STANDARD; PRT; 769 AA.
 AC P38920;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE MUTL protein homolog 1 (DNA mismatch repair protein MLH1).
 GN MLH1 OR YMR167M OR YMR520.16.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RP [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9408838; PubMed=8264608;
 RA Prolia T.A., Christie D.-M., Liskay R.M.;
 RT "Dual requirement in yeast DNA mismatch repair for MLH1 and PMS1, two
 RT homologs of the bacterial mutL gene.";
 RL Mol. Cell. Biol. 14:407-415(1994).
 RL [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C; AB972;
 RC MEDLINE=97313268; PubMed=9169872;
 RA Bowman S., Churcher C.M., Badcock K., Brown D., Chillingworth T.,
 RA Connor R., Dedman K., Devlin K., Gentles S., Hamlin N., Hunt S.,
 RA Jagers K., Lye G., Moule S., Odell C., Pearson D., Raglandream M.A.,
 RA Rice P., Skelton J., Walsh S., Whitehead S., Barrell B.G.;
 RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome
 RT XII.";
 RL Nature 387:90-93(1997).
 RN [3]
 RN MUTAGENESIS OF ALA-41; GLY-64; ILE-65; GLU-99; ILE-104; THR-114;
 RP ARG-214; VAL-216; ARG-265; ILE-326; GLN-552; ARG-672 AND ALA-694.
 RX MEDLINE=21439334; PubMed=11555625;
 RA Ellison A.R., Lofing J., Bitter G.A.;
 RT "Functional analysis of human MLH1 and MSH2 missense variants and
 RT hybrid human-yeast MLH1 proteins in Saccharomyces cerevisiae.";
 RL Hum. Mol. Genet. 10:1889-1900(2001).
 CC -1- FUNCTION: This protein is involved in the repair of mismatches
 CC in DNA. Plays a major role in maintaining the genetic stability
 CC of simple sequence repeats and in the repair of heteroduplex
 CC sites present in meiotic recombination intermediates. Probably
 CC MSH2, PMS1, and MLH1 are components of the same DNA mismatch
 CC repair pathway.
 CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -1- SIMILARITY: Belongs to the DNA mismatch repair mutL/hexB family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U07187; AAA16835.1; -;
 DR EMBL; 249705; CAA899803.1; -;
 DR PIR; S54525; S54525.
 DR HSP; P23367; IERN.
 DR GenBank; 142839; -;
 DR SCD; S0004777; MLH1.
 DR GO; GO:0005634; C:nucleus; IMP.
 DR GO; GO:0003677; F:DNA binding; IDA.

```

DR InterPro: IPR003594: Atphind_Atpase.
DR InterPro: IPR002099: DNA_mis_repair.
DR Pfam: PF01119: DNA_mis_repair; 1.
DR Pfam: PF02518: HATPase_C; 1.
DR SMART: SM00367: HATPase_C; 1.
DR TIGRFAMs: TIGR00585; multi; 1.
DR PROSITE: PS00058: DNA_MISMATCH_REPAIR_1; 1.
DR DNA_repair: Nuclear protein.
KV MUTAGEN 41 41 A->F: DEFECTIVE IN A MISMATCH REPAIR
FT MUTAGEN 41 41 ASSAY.
FT MUTAGEN 41 41 A->S: FULLY FUNCTIONAL IN A MISMATCH
FT MUTAGEN 64 64 REPAIR ASSAY.
FT MUTAGEN 64 64 G->R: DEFECTIVE IN A MISMATCH REPAIR
FT MUTAGEN 65 65 ASSAY.
FT MUTAGEN 65 65 I->N: DEFECTIVE IN A MISMATCH REPAIR
FT MUTAGEN 99 99 ASSAY.
FT MUTAGEN 99 99 E->K: DEFECTIVE IN A MISMATCH REPAIR
FT MUTAGEN 104 104 ASSAY.
FT MUTAGEN 104 104 I->R: DEFECTIVE IN A MISMATCH REPAIR
FT MUTAGEN 114 114 T->R: DEFECTIVE IN A MISMATCH REPAIR
FT MUTAGEN 214 214 ASSAY.
FT MUTAGEN 214 214 R->C: PARTIALLY DEFECTIVE IN A MISMATCH
FT MUTAGEN 216 216 REPAIR ASSAY.
FT MUTAGEN 216 216 V->I: FULLY FUNCTIONAL IN A MISMATCH
FT MUTAGEN 265 265 REPAIR ASSAY.
FT MUTAGEN 265 265 R->C: PARTIALLY DEFECTIVE IN A MISMATCH
FT MUTAGEN 265 265 REPAIR ASSAY.
FT MUTAGEN 265 265 R->H: PARTIALLY DEFECTIVE IN A MISMATCH
FT MUTAGEN 326 326 REPAIR ASSAY.
FT MUTAGEN 326 326 I->A: PARTIALLY DEFECTIVE IN A MISMATCH
FT MUTAGEN 326 326 I->V: FULLY FUNCTIONAL IN A MISMATCH
FT MUTAGEN 552 552 REPAIR ASSAY.
FT MUTAGEN 552 552 O->L: DEFECTIVE IN A MISMATCH REPAIR
FT MUTAGEN 672 672 ASSAY.
FT MUTAGEN 672 672 R->P: DEFECTIVE IN A MISMATCH REPAIR
FT MUTAGEN 694 694 ASSAY.
FT MUTAGEN 694 694 A->T: FULLY FUNCTIONAL IN A MISMATCH
FT CONFLICT 258 258 P->L (IN REF. 1).
FT CONFLICT 288 288 N->F (IN REF. 1).
FT CONFLICT 708 708 S->L (IN REF. 1).
SQ SEQUENCE 769 AA; 87062 MW; B2DBB31DB3943171 CRC64;

```

Query Match 6.2%; Score 110; DB 1; Length 769;

Best Local Similarity 20.1%; Pred. No. 4.2; Matches 63; Conservative 58; Mismatches 125; Indels 68; Gaps 11;

```

QY 13 NENGS---VDIAREMILKAKAGNNAVFCFKDKLISALAPRAVQIKNTGLESQ 68
DB 165 SHADYSKLDVAGYVAISHKDIGSCCKF---GDSNYS-LSVPSSTVQ----- 210
QY 69 EMTKLEMKYDDYLHEKEVAAGSLNDFSTPDED---SIDFLASLKQIKWIPSGEL 124
DB 211 -----DRIRIVNKSVAASNLITFIHSKVEDNLNESVD-----GRVC 246
QY 125 NIPLYEKAKLPIDPKKIIISTGMATIDIKOSVSIFINNKIPVGNITITHCETIYP 184
DB 247 NINFTSKSKISPI---PFINNRLVYCDILIRALMSVSNVLPKKNRPETIYGIVIDPA 302
QY 186 EDVNIANINIDLKHPKKNIGFSDH-----SSGFYAIAPVGIPTIEGFTIDKSM 237
DB 303 VDYNV-----HPTKREVRFLSDDEITIKIANQHLALSILDSRTTKASISITNKE 354
QY 238 SGPDLIASIEPDELKALCIQVRCVEKS-LGSNSKVTFTASERKOKIYAKSIIAK--TRIK 294
DB 355 SLIPFDLTIESDRNRKSLROAQVENSYTLANSOLAKAKROENKIVRIDASQAKITSFIS 414
QY 295 KGBVSEKNIITTK 308
DB 415 SSGQFNFECSSTK 428

```

RESULT 6

```

ID ISPG_CHLMU STANDARD; PRT; 601 AA.
AC Q9PKT3;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE 4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase (EC 1.17.4.3) (1-
DE hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase).
GN ISPG OR TC0327.
OS Chlamydia muridarum.
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83560;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Mopn / N159;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Ultezhack T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwyn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia
RT pneumoniae A39."
RL Nucleic Acids Res. 28:1397-1406(2000).
CC -I- FUNCTION: Converts 2C-methyl-D-erythritol 2,4-cyclodiphosphate
CC (ME-2,4cp) into 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate
CC (By similarity).
CC -I- CATALYTIC ACTIVITY: (E)-4-hydroxy-3-methylbut-2-en-1-yl
CC diphosphate + H(2)O + protein-dithiol = 2-C-methyl-D-erythritol
CC 2,4-cyclodiphosphate + protein-dithiol.
CC -I- PATHWAY: Nonaevulonate terpenoid biosynthesis pathway; sixth step.
CC -I- SIMILARITY: Belongs to the ispg family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation at
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AE002300; AAP9291.1; -
CC PIR: C81715; C81715.
DR TIGR; TC0327; -
DR HAMAP; MF_00159; -; 1.
DR InterPro: IPR004588; ispg.
DR Pfam; PF04551; Gcpg; 1.
DR TIGRFAMs; TIGR00612; ispg_gcpg; 1.
KW Isoprene biosynthesis; Complete proteome; Oxidoreductase.
SQ SEQUENCE 601 AA; 67106 MW; CFA673DB9DD1E2B4 CRC64;

```

Query Match 6.1%; Score 109; DB 1; Length 601;

Best Local Similarity 20.0%; Pred. No. 3.6; Matches 81; Conservative 57; Mismatches 122; Indels 146; Gaps 19;

```

QY 4 IYIABIGCNGNSVDIAREMILKAKAGV-NAVFCFKDKLISALIA-----PKAEYQ 57
DB 50 IYAIQECG-----EIVRTVQGLKEVAGCEQKIDRLVQGNVITPLVADHFFPQAIIH 103
QY 58 IKN-----TSELPSQLEMTKLEKQDDYHL----- 84
DB 104 VADFDKVRIRPGNVYVDKRNPFSGKIYSDDEQYTSRLRPFKFSPLVAKCRILGRARIG 163
QY 85 -----MEXA-VSKLVDPSTPFDSDSID-----FL 108
DB 164 VNHGSLSERINQRYGDTIEGVAFSALEAYACVMDYKNIYFSSKSSNPRMVAAYRLA 223
QY 109 ASLKQIKWIPSGELANPLYEKIAPIDPKKIIISTGMATIDIKOSVSIFIT 162
DB 224 RELDQRKVIY-----LHLGVTEAGSGM---GGMKISVIGICTLSEGLDITRCSITGSP 276

```

QY 163 NKKVAVGNITILECNTYEP-----TPEDVNLNAI--NDLKHHPKXNIGSDHSGCY 214
 DB 277 TLEIIP-CDLKLKTKAKSKTKKVPFIYHSKULTTQTRKHPPVENV-----YGLT 329
 QY 215 AIAIAVPYGTTFEKAFTLDDKSMGPDHLASIEPDELKLCIGVR-----C 260
 DB 330 V-----GHT-----KDHLLTEPNTLD-CISVDITTKKDLTPDGVV 367
 QY 261 VEKLSGSKVATASERKXKIVAKSIITAKTEIKKGEVSEKNIT 306
 DB 368 IKSMS-SAISVEIKHARIPYKEDAPILNENNELMISDITLAT 412
 RESULT 7
 YG38 YEAST STANDARD; PRT; 713 AA.
 AC P53276;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Hypothetical 80.2 kDa protein in ASN2-PH1 intergenic region.
 GN YGR128C.
 OS Saccharomyces cerevisiae (Baker's Yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA van Dyck L., Skala J., de Wergifosse P., Purnelle B., Talla B.,
 RA Nawrocki A., del Bino S., Goffeau A.;
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; Z72913; CA97140.1; -
 DR PIR; S64437; S64437.
 DR Germonline; 141440; -
 DR SGD; S0003360; YGR128C.
 DR GO; GO:0005132; C:small nucleolar ribonucleoprotein complex; IPI.
 DR GO; GO:0030515; R:snRNP binding; IPI.
 DR GO; GO:0030490; P:processing of 20S pre-rRNA; IMP.
 DR KW Hypothetical protein; Transmembrane.
 FT TRANSMEM 635 655 POTENTIAL.
 FT SEQUENCE 713 AA; 80190 MW; 655AE1492F841E49 CRC64;
 Query Match 6.0%; Score 107.5; DB 1; Length 713;
 Best Local Similarity 19.5%; Pred. No. 5.6; Indels 115; Gaps 18;
 Matches 76; Conservative 69; Mismatches 130;
 QY 14 HNGSVDIAREMILKARAGAVNAVFOTFKADKLISAIAPRAEYQIKNGELSSQLEMTTK 73
 DB 157 YTSDEFLSQQYKIKAK-AKWSIKIDT--KNSLIVAILQNGLIEI-----FPKLTILHS 208
 QY 74 LEKQYDYLAHMEYANSINDVF-----STPFDSDIDFLA 109
 DB 209 FDISYDNLKAKAFTEGTEYVVLCPLODDKVCYKLELTDCSGSSSSPIYELSTIIE 268
 QY 110 SLK-----QIKWIKPSGE--LILNIFY--LEKIKALPIPKKIIISTGMATIDEIK 155
 DB 269 GPFSEFKCYGPGKIDYKLNQKIKYIYSAPHQLOQYIEPMDK--LSPGDILISFOP 325
 QY 156 QASV---IFLNKVPVGNITILKNT-----EYPPFEVNLNAINDLKHPKXNIGF 206
 DB 326 VSVNRLVLTANNVIYL--LDLHCSTLSQELTHVKTFOILKSAVINSEKSHSK----- 378
 QY 207 SDHSSGFYALIAVPYGTTFEKAFTLDDKSMGPDHLASIEPDELKLCIGVRCEVSKJG 266

DB 379 -----TIAIGIS-----TNGNPFSSL--EIIINDVCTNLTQSLG 413
 QY 267 SNSKVATASERKXKIVAKSIITAKTEIKKGEVSEKNITTK-----PENGISPM----- 316
 DB 414 KSPQVGNND-----SSVILK-----ELPDDKDIINDRKVKNVDVSGSSVPVLAHCN 458
 QY 317 EYNILKIAADPPIPELIIHSEPKNGE 346
 DB 459 EVIERLSALQNDITSPDIPFKELKIKEE 488
 RESULT 8
 NDM1 YEAST STANDARD; PRT; 2748 AA.
 ID NDM1_YEAST
 AC Q00402; Q03767;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Nuclear migration protein NDM1.
 GN NDM1 OR YDR150W OR YD8358.06.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A. / FL100;
 RC STRAIN=ATCC 28363 / FL100;
 RX MEDLINE=92079907; PubMed=1145235;
 RA Kornanev J., Schaefli-Geertsenschlaeger I., Zimmermann F.K.,
 RA Perecko D., Kuentzel H.;
 RT "Nuclear migration in 313 kDa NDM1 protein";
 RL Mol. Gen. Genet. 230:277-287(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SS288C / FY1678;
 RX MEDLINE=97313263; PubMed=9169667;
 RA Jacq C., Alt-Moerbe J., Andre B., Arnold W., Bahr A., Ballestra J.P.G.,
 RA Barques M., Baron L., Becker A., Bileau N., Bloecker H., Blugnon C.,
 RA Boskovic J., Brandt P., Bruckner M., Buitrago M.J., Coster F.,
 RA Delaveau T., del Rey F., Dujon B., Ride L.G., Garcia-Camalejo J.M.,
 RA Goffeau A., Gomez-Pertis A., Granclier C., Henemann V., Hankeln T.,
 RA Hottelisel J.D., Jaeger W., Jimenez A., Jomiaux J.-L., Kremer C.,
 RA Kuester H., Laamann P., Legros Y., Louis E.J., Koeller-Elexer S.,
 RA Monnet A., Moro M., Mueller-Auer S., Nussbaumer B., Paricio N.,
 RA Paulin L., Pera J., Perez-Alonso M., Perez-Ortin J.E., Pohl T.M.,
 RA Rieger W., Purnelle B., Rasmussen S.W., Remacha M., Revuelta J.L.,
 RA Scharte M., Salom D., Saluz H.P., Sait J.E., Saren A.-M., Schaefer M.,
 RA Scharte M., Schmidt E.R., Schneider C., Scholler P., Schwarz S.,
 RA Uriestaran L.A., Verhaesele P., Vissers S., Voet M., Voickert G.,
 RA Wagner G., Wambutt R., Wedler S., Wedler S., Weill S., Harris D.E.,
 RA Bowman S., Brown D., Churcher C.M., Connor R., Dedman K., Gentiles S.,
 RA Hamlin N., Hunt S., Jones L., McDonald S., Murphy L., Niblett D.,
 RA Odell C., Oliver K., Rajandream M.A., Richards C., Shore L.,
 RA Walsh S.V., Barrall B.G., Dietrich F.S., Mulligan J.T., Allen E.,
 RA Araujo R., Aviles E., Bero A., Carpenter J., Chen E., Cherry J.M.,
 RA Chung E., Duncan M., Hunkle-Smith S., Hyman R.W., Kemp C.,
 RA Lasakati D., Lew H., Lin D., Mosedale D., Nakatara K., Natch A.,
 RA Oehler P., Oh C., Petel F.X., Roberts D., Schramm S., Schroeder M.,
 RA Shogren T., Shroff N., Winant A., Yellon M.A., Botstein D.,
 RA Davis R.W., Johnston M., Andrews S., Brinkman R., Cooper J., Ding H.,
 RA Du Z., Pavllo A., Fulton L., Gattung S., Greco T., Hallsworth K.,
 RA Hacking J., Hillier L.W., Jier M., Johnson D., Johnson L.,
 RA Kirzen S., Kocaba T., Langston Y., Latreille P., Le T., Mardis E.,
 RA Meneses S., Miller N., Nhan M., Pauley A., Peluso D., Ritkin L.,
 RA Riles L., Tatch A., Trevarakis E., Vignati D., Wilcox L., Wohlmann P.,
 RA Vaudin M., Wilson R., Westerton R., Albertmann K., Hand J., Heumann K.,
 RA Kleene K., Mewes H.-W., Zollner A., Zaccaria P.;
 RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome IV";
 RL Nature 367:75-78(1997).
 CC -I- FUNCTION: Controls nuclear migration. NDM1 specifically controls
 CC the interaction of the bud neck cytoskeleton with the pre-
 CC divisional G2 nucleus perhaps by recognizing G2-specific

cytoplasmic microtubuli or other components of the nuclear envelope.

CC MISCELLANEOUS: Additional regions of lower homology to the repeat consensus (always starting with proline) are found in both flanking domains of the tandem repeats.

CC -1- SIMILARITY: Contains 1 PH domain.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).

CC -----

CC EMBL: X61236; CAA43554.1; .

CC EMBL: Z50046; CAA90372.1; .

CC GenBank: 140641; .

CC SGI: S0002557; NMI.

CC GO: GO:0005934; C:bud tip; IDA.

CC GO: GO:0005938; C:cell cortex; IDA.

CC GO: GO:0015631; F:tubulin binding; IPI.

CC GO: GO:0000226; P:microtubule cytoskeleton organization and b. . . ; IMP.

CC GO: GO:0000065; P:nuclear migration (sensu Saccharomyces); IMP.

CC InterPro: IPR005529; DUF321.

CC InterPro: IPR001849; PH.

CC Pfam: PF03778; DUF321; 13.

CC Pfam: PF00169; PH; 1.

CC SMART: SM00233; PH; 1.

CC PROSITE: PS00003; PH DOMAIN; 1.

CC -----

CC Repeat:

CC DOMAIN 593 1384 12.5 X TANDDEM REPEATS.

CC REPEAT 593 656 1.

CC REPEAT 657 727 2.

CC REPEAT 728 798 3.

CC REPEAT 799 862 4.

CC REPEAT 863 926 5.

CC REPEAT 927 990 6.

CC REPEAT 991 1054 7.

CC REPEAT 1055 1118 8.

CC REPEAT 1119 1182 9.

CC REPEAT 1183 1246 10.

CC REPEAT 1247 1310 11.

CC REPEAT 1311 1374 12.

CC REPEAT 1375 1384 13 (INCOMPLETE).

CC DOMAIN 2573 2683 PH.

CC CONFLICT 1570 1570 A->V (IN REF. 1).

CC CONFLICT 1822 1822 E->K (IN REF. 1).

CC CONFLICT 1960 1962 KAS->RH (IN REF. 1).

CC CONFLICT 1971 1972 RD->RN (IN REF. 1).

CC CONFLICT 2049 2049 S->N (IN REF. 1).

CC CONFLICT 2637 2637 V->A (IN REF. 1).

CC SEQUENCE 2748 AA; 313030 MW; EBA448E950621142 CRC64;

Query Match 6.0%; Score 107.5; DB 1; Length 2748;

Best Local Similarity 21.1%; Pred. No. 30;

Matches 80; Conservative 61; Mismatches 113; Indels 125; Gaps 19;

CC 18 VDIARE-----MILAKKAGVAVVQCFKADKLISALAPKAYQIKNGELBSQLEMTKK 73

CC 1871 VAISQDYELVNAKREKPOLYTLKISSKIDHT--VPLSEVNLVNTNRPSELSTLKE 1927

CC 74 -----LEMKYDDY-----LHLMKAVVSLN-----LDVEST--PDEDSD 106

CC 1928 KAVLNHLLIKEDDYKNILAVSEHPVILHSEKASLKNKVLVDODFATMSRSTLEKPTID 1987

CC 107 FLASIKQKIKWIPSGE-----LHNPYLEKAKLPPIPKKIIS-----TKM 148

CC 1988 FLSTALSMGKILVNESTHKNEKLSPEDEPLTKAKKQGLIISKEYSELRDQIDR 2047

CC 149 AIIIEIKQSVSIFINNVKPVGNITILH-CNTEVEPT--FEDVNLNAINDLKKHFPKKN 204

CC 2048 PSLDTLKKKAIF--DSITIVENIEQGLVNTTSCPPITTEDLKVYA----- 2092

CC 205 GFDSSSGFYAAIAVAYGITFIKFFTLDSKSGPDLASIEPDELKHLICIVRCYKS 264

CC 2093 -----HKGMELCLOKPKKSGARAE-----RIDEQS 2120

CC 265 LG--SNKVVVTA-----ERKKIVARXS-----IIKTE-IKKGEV--SRKNI 304

CC 2121 INTTSSSTTSSKFTALDNDNIELNRVELONNEDYTIIKSTVYDITIFPAVENI 2180

CC 305 --TTKRKNGISPEWNTL 321

CC 2181 KNSAEKIGYLVPEKSN 2195

CC -----

CC RESULT 9

CC YOPM_YERPE STANDARD; PRT; 409 AA.

CC ID YOPM_YERPE

CC AC P1778; O68701;

CC DT 01-AUG-1990 (Rel. 15, Created)

CC DT 28-FEB-2003 (Rel. 41, Last sequence update)

CC DT 10-OCT-2003 (Rel. 42, Last annotation update)

CC DE Outer membrane protein YOPM.

CC GN YOPM OR YOP48 OR YPCD1.26C OR Y5054 OR Y0059.

CC OS Yersinia pestis.

CC OG Plasmid pCD1.

CC OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

CC OC Enterobacteriaceae; Yersinia.

CC OX NCBI_TaxID=632;

CC [1]

CC RP SEQUENCE FROM N.A.

CC RC STRAIN=KIMS / Biovar Mediaevalis;

CC RX MEDLINE=89359090; PubMed=2670888;

CC RA Leung K.Y., Straley S.C.;

CC RT "The yopM gene of Yersinia pestis encodes a released protein having homology with the human platelet surface protein GpIIb alpha.";

CC RL J. Bacteriol. 171:4623-4632(1989).

CC [2]

CC RP SEQUENCE FROM N.A.

CC RC STRAIN=KIMS / Biovar Mediaevalis;

CC RX MEDLINE=98427122; PubMed=9746557;

CC RA Perry R.D., Straley S.C., Fetherston J.D., Rose D.J., Gregor J., Blatner F.R.;

CC RT "DNA sequencing and analysis of the low-Ca2+-response plasmid pCD1 of Yersinia pestis KIMS.";

CC RL Infect. Immun. 66:4611-4623(1998).

CC [3]

CC RP SEQUENCE FROM N.A.

CC RC STRAIN=KIMS / Biovar Mediaevalis;

CC RX MEDLINE=98422474; PubMed=9748454;

CC RA Hu P., Elliott J., McCreedy P., Skowronski E., Garmes J., Kobayashi A., Brubaker R.R., Garcia B.;

CC RT "Structural organization of virulence-associated plasmids of Yersinia pestis.";

CC RL J. Bacteriol. 180:5192-5202(1998).

CC [4]

CC RP SEQUENCE FROM N.A.

CC RC STRAIN=CO-92 / Biovar Orientalis;

CC RX MEDLINE=21470413; PubMed=11586360;

CC RA Partholl J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G., Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L., Baser S., Basham D., Bentley S.D., Brooks K., Cedeno-Tarrega A.M., Chillingworth T., Cronin A., Davies R.M., Davis F., Dougan A.V., Fellwell T., Hamlin N., Holroyd S., Jagels K., Kariyasev A.V., Leachon S., Moulton S., Oyston P.C.F., Quail M.A., Rutherford K., Simmonds S., Skelton J., Stevens K., Whitehead S., Barrett B.G.;

CC RT "Genome sequence of Yersinia pestis, the causative agent of plague.";

CC RL Nature 413:523-527(2001).

CC CC -1- STRUCTURAL LOCATION: OUTER MEMBRANE AND SECRETED.

CC CC -1- SIMILARITY: Contains 12 leucine-rich (LRR) repeats.

CC CC -1- SIMILARITY: TO S.FLEXNERI IPAN 7.8.

CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; AF074612; AAC69806.1; -
 DR EMBL; AF053946; AAC62580.1; -
 DR EMBL; AL117189; CAB54903.1; -
 DR PIR; A33950; A33950.
 DR PIR; T43599; T43599.
 DR InterPro; IPR001611; LRR.
 DR InterPro; IPR007088; LRR_Dact.
 DR Pfam; PF00560; LRR_7.
 KM Outer membrane; Repeat; leucine-rich repeat; Antigen; Plasmid;
 KM Complete proteome.
 FT REPEAT 69 90 LRR 1.
 FT REPEAT 91 110 LRR 2.
 FT REPEAT 111 129 LRR 3.
 FT REPEAT 130 152 LRR 4.
 FT REPEAT 153 177 LRR 5.
 FT REPEAT 179 194 LRR 6.
 FT REPEAT 195 217 LRR 7.
 FT REPEAT 236 256 LRR 8.
 FT REPEAT 257 281 LRR 9.
 FT REPEAT 296 316 LRR 10.
 FT REPEAT 317 341 LRR 11.
 FT REPEAT 356 381 LRR 12.
 FT CONFLICT 142 183 MISSING (IN REF. 1).
 SQ SEQUENCE 409 AA; 46203 MW; E22400C671A59951 CRC64;

Query Match 6.0%; Score 107; DB 1; Length 409;
 Best Local Similarity 20.3%; Pred. No. 3.1; 97; Indels 48; Gaps 12;
 Matches 53; Conservative 63; Mismatches 97; Indels 48; Gaps 12;

CC 45 KLISALPKAEYQIKNTGELSOLEMTKLEMYDYILHMEYAVSLNDFSTPDED 104
 DB 124 KALSDLEPLEY---LGVSNNGLE--KLPELGNSSFLKIIDVD--NNSLKLPLPLPS 174
 QY 105 IDPLASLKQKIMIPSELINLY-----LEKIALPLPDKKIISTGMATIDEI 154
 DB 175 LELIAAGNNQLELP--ELQNLPELTAIYADNNSLKLPLPLPSIY--AGNLIIEH 230
 QY 155 KQVSIFINKKVPVGNITLHCTEYPTPEVDNL--NAINDLKKAEPKKNIGFSDHSG 212
 DB 231 PELQNLPELTTIYADN--NLKTLPELPPSLHAIYADNNTLDFE--LPO--SLTFIDVSEN 287
 QY 213 FYAIAAVPYGIFTEKHFPLDKMSPDHASIEPDELKHLCTIGVCVEKSTGNSKYV 272
 DB 288 IFSGLSLEPLNLYL-----NASSNEIRSLCDPLPSLEHIVNSNKKLI 330
 QY 273 ---TASERKKIVARKSIIX 290
 DB 331 ELPALEPRLERLISFNLAE 351

RESULT 10
 YA44 SCHPO STANDARD; PRT; 1428 AA.
 ID YA44 SCHPO
 AC 009773;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Hypothetical protein C22F3.04 in chromosome I.
 GN SPAC22F3.04.
 OS Schizosaccharomyces pombe (Fission Yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomyces.
 NCBI_TaxID=4896;
 RX NCBI_TaxID=4896;
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;

RX MEDLINE=21848401; PubMed=11859360;
 RA Wood V., Galliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Church C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
 RA James K., Jones L., Jones M., Leacher S., McDonald S., McLean J.,
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinovitch E.,
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
 RA Taylor K., Taylor R.G., Tiley A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Volkart G., Aert R., Robben D., Grympey B.,
 RA Wellens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Pritzc C., Holzer E., Moesti D., Hilbert H.,
 RA Borzym K., Langer I., Beck A., Leinrich R., Reinhardt R., Pohl T.M.,
 RA Eger P., Zimmermann W., Medler H., Mambutt R., Purnelle B.,
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaue V., Motier S.,
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas W., Rochet M., Gallard C., Tallada V.A., Garzon A., Thode G.,
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
 RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 RA Shpakovski G.V., Useery D., Barrett B.G., Nurse P.,
 RT "The genome sequence of Schizosaccharomyces pombe".
 RL Nature 415:871-880(2002).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (potential).
 CC -1- SIMILARITY: TO S.POMBE SPAC56P8.02.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; 254285; CAB91069.1; -
 DR PIR; T38187; S62419.
 DR GeneDB; SPAC22F3.04; -
 DR InterPro; IPR000873; AMP-bind.
 DR Pfam; PF00501; AMP-binding; 3.
 KM Hypothetical protein; Transmembrane.
 FT TRANSMEM 1028 1048 POTENTIAL.
 FT TRANSMEM 1303 1323 POTENTIAL.
 FT TRANSMEM 1333 1353 POTENTIAL.
 SQ SEQUENCE 1428 AA; 162394 MW; A88D03AE0A9E9E99 CRC64;

Query Match 5.9%; Score 105.5; DB 1; Length 1428;
 Best Local Similarity 23.5%; Pred. No. 18;
 Matches 79; Conservative 44; Mismatches 144; Indels 69; Gaps 14;

CC 14 HNGSYDIARENT-----LKAKEAGV--NAVYGTETKADKLISALPKAEYQIKNTGE 63
 DB 785 HONKVSQKARENLMHFTLLDPLKTSKAPTSPALITFNALTEKVELTMAHFEKLVASH 844
 QY 64 LE-SQLENTYK---LEMKYDDYLHMEYAVSLNDFSTPDEDSDIDPLASLKQKIMK 117
 DB 845 VEYQGLSVAKARSHILLVYDPLEFLSIHSCFLVGIPIPCFQIREISQILGEIEFTR 904
 QY 118 IPSC-----ELNLP-----YLEKIALPLPDKKIISTGMATIDEIKQVS 159
 DB 905 IAKAPVZAILVDHVKVLSIKSRDISNHFQCTCIDLNKAPRIEFTTIDIPISKRAKALN 964
 QY 160 IF-----INNKVPVGNITL--LHCTEYPTPEVDNLAINDLKKAEPKKNIGFSD 208
 DB 965 LITPELANKKQKVALISINKLEDSIIPTQSHSLAFCHQKEFVIGNEKPIIGIE 1024
 QY 209 HSSG---FYAIAAVPYGI--TFIEKHFPLDKMSPDHASIEPDELKHLCTIGVCVEKS 264
 DB 1025 FSSQWGLHTALTALGVYAVPTLLIKQENTL--CNNGSILFRAIEONSLSKVMIPLNT 1082

QY 265 L-----GNSKRVTA-----SEKKNIV 282
DB 1083 FSTAGGNCIAIVNSTESSIVPCYDRPISSRVNII 1118

RESULT 11
ID YFE9_YEAST STANDARD; PRT; 623 AA.
AC P43554;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Hypothetical 70.3 kDa protein in ALK2-EMP4 intergenic region.
GN YF049W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=5288C / AB972;
RX MEDLINE=99400292; PubMed=7670463;
RA Murakami Y., Naitou M., Hagiwara H., Shibata T., Ozawa M.,
RA Sasanuma S.-I., Sasanuma M., Tsuchiya Y., Soeda E., Yokoyama K.,
RA Yamazaki M., Tashiro H., Eki T.;
RT Analysis of the nucleotide sequence of chromosome VI from
RT Saccharomyces cerevisiae.
RL Nat. Genet. 10:261-268(1995).
CC -1- SIMILARITY: TO YEAST NPL6.
CC -1-
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: D50617; BAA09192.1; -
DR PIR: S56206; S56206.
DR GeneOnline: 140107;
DR SGD: S0001845; YFL049W.
DR Hypothetical protein.
SQ SEQUENCE 623 AA; 70275 MW; A8AC00CA8F0ED0E CRC64;

Query Match 5.8%; Score 103.5; DB 1; Length 623;
Best Local Similarity 22.3%; Pred. No. 8.9;
Matches 86; Conservative 41; Mismatches 134; Indels 125; Gaps 19;

QY 9 EIGNGNGSVDIARE-----MILKKEAGVNAVKQOTFKADKLISALPKQEVQIKNT 61
DB 254 EISANTVNTCTISKFSQSPYVTEQPSAIREAVLENFAGKSHISAIVP--GQSIGT 311

QY 62 CELESQLEMKKLEMKVDVILHKEVAVSLNDVFSPPEDSDIDELASIKQIKWIPSG 121
DB 312 LEISAQ-----FVPRYHKNKSFQALQAMANDPIG 343

QY 122 ELNMLPYLEKIAKL--PIPKKIIISGMATIDIKOSVIFINKKPVQNT--TILHGN 177
DB 344 R-----HEBLLAQESQAPD-----GSAISISLPHNIPSVNPK--PIKRMSSILDIN 390

QY 178 TEVTPPEDEVNAIINLKLKHPKQNGF-----SHSGG--FYAIAAIP 221
DB 391 VSSSKNKSEENEMIKWNGKQKHNNVSLNGKFSLSPLKSLKNSGKQOYR----- 444

QY 222 YGIFIKKFLTIDSMGSPDHASIEPDELK--HLCIGVCYKESKSGNSKVT----- 273
DB 445 -GLVLYEKNTLLER-----LKQLTNEIKELHLDAY--FVVTGLOVNRKRTKKKK 495

QY 274 -----ASERNKIVAKSIATKTEIKGVSFSE-----KNITTKRGNG 312
DB 496 YWQYKAGIPIGLKRSQLEDFKRYL--KDVLAQSVATN--FNEITWDTBITTKRVNPN 551

QY 313 ISPMENYLLGKIAEDPIDELIIH 338
DB 552 -----NEFGNCHIDFKPPIYSH 570

RESULT 12
ID PYCB_METUA STANDARD; PRT; 567 AA.
AC O58628;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Pyruvate carboxylase subunit B (EC 6.4.1.1) (Pyruvic carboxylase B).
GN PYCB OR MJ1231.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kierulff A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.L.,
RA Overbeek R., Kirkness E.F., Weissbrock K.G., Merrick J.M., Nguyen D.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhmann J.L., Hanna M.C.,
RA Uitterlinden L.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurt W.A., Kane B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii".
RL Science 273:1058-1073(1996).
RN [2]
RP SEQUENCE OF 190-125; 260-270; 277-289; 277-289; 309-325; 328-358;
RP 370-380; 386-409; 422-438; 491-506 AND 491-506, AND FUNCTION.
RX MEDLINE=2104791; PubMed=1195096;
RA Mukhopadhyay B., Patel V.J., Wolfe R.S.;
RT "A stable archaeal pyruvate carboxylase from the hyperthermophile
RT Methanococcus jannaschii".
RL Arch. Microbiol. 174:406-414(2000).
CC -1- FUNCTION: Pyruvate carboxylase catalyzes a 2-step reaction,
CC involving the ATP-dependent carboxylation of the covalently
CC attached biotin in the first step and the transfer of the carboxyl
CC group to pyruvate in the second.
CC -1- CATALYTIC ACTIVITY: ATP + pyruvate + HCO(3) (-) = ADP + phosphate +
CC oxaloacetate.
CC -1- COPROCTOR: ATP, magnesium (or manganese or cobalt), pyruvate and
CC bicarbonate.
CC -1- ENZYME REGULATION: INHIBITED BY MAGNESIUM, WHEN ITS CONCENTRATION
CC EXCEEDED THE ATP ONE, AND BY HIGH CONCENTRATION OF ATP AND ALPHA-
CC KETOGLUTARATE.
CC -1- PATHWAY: Gluconeogenesis.
CC -1- SUBUNIT: Heterooctamer of four A and four B subunits.
CC -1- MASS SPECTROMETRY: MW=64160; METHOD=VALDIT.
CC -1- WISCELLANEOUS: ITS OPTIMUM PH IS 8.5 AND THE OPTIMUM TEMPERATURE
CC IS 80-90 DEGREES CELSIUS.
CC -1- SIMILARITY: WITH OTHER BIOTIN CARBOXYLASES, LIPOMIDE TRANSFERASES
CC AND CARBAMYL PHOSPHATE SYNTHETASES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U67563; AAB99233.1; -
DR PIR: P64453; P64453.
DR HSSP: P02905; 1A6X.
DR TIGR: MJ1231; -
DR InterPro: IPR001882; Biotin_BS.

Thu May 13 11:53:14 2004

us-09-930-440b-8.rsp

Page 2

Db 1 MSNIYVAEIGCHNGSVDIAREMILKAKAGVNAVFOETKADKLISAPRAEYOJIN 60
Qy 61 TGELESQLEMTKLEMKYDDYHMEYAVSLNDVSTPDESDIDFLASKOKIMKIPS 120
Db 61 TGELESQLEMTKLEMKYDDYHMEYAVSLNDVSTPDESDIDFLASKOKIMKIPS 120
Qy 121 GELINLPYLEKIAKLPIPDCKIIISTGMATIDEIKOSVSIFINNKVAVGNTITLHCNTY 180
Db 121 GELINLPYLEKIAKLPIPDCKIIISTGMATIDEIKOSVSIFINNKVAVGNTITLHCNTY 180
Qy 181 PTFEPEDVNAINDLKHPKNNIGFSDHSGFYAAIAAPVIGITTEKFTLTKMSGP 240
Db 181 PTFEPEDVNAINDLKHPKNNIGFSDHSGFYAAIAAPVIGITTEKFTLTKMSGP 240
Qy 241 DHIASIEPDELKHLICIGVRCVKSLSGNSKVVTASERKNKIVAKSIIAKTEIKGEVFS 300
Db 241 DHIASIEPDELKHLICIGVRCVKSLSGNSKVVTASERKNKIVAKSIIAKTEIKGEVFS 300
Qy 301 EKNTTRPGNGISPMEMWNLGKIAEODFIDELIHSEFNOE 346
Db 301 EKNTTRPGNGISPMEMWNLGKIAEODFIDELIHSEFNOE 346

RESULT 2

Q93NQ2 PRELIMINARY; PRT; 346 AA.
AC Q93NQ2 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE NAB.
GN Escherichia coli.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN NCBI_TaxID=562;
RP SEQUENCE FROM N.A.
RX MEDLINE=2129197; PubMed=11404020;
RA Wang L., Briggs C.E., Rothmund D., Fratamico P., Luchansky J.B.,
Reeves P.R.;
RT "Sequence of the E. coli O104 antigen gene cluster and identification
of O104 specific genes.";
RL Gene 270:231-236(2001).
DR EMBL; AF61371; AAK64368.1; -;
DR GO; GO:0016051; P:carbohydrate biosynthesis; IEA.
DR InterPro; IPR006014; Antifreeze dom.
DR InterPro; IPR006190; Antifreeze-like.
DR Pfam; PF01354; Antifreeze; 1.
DR Pfam; PF01354; Antifreeze; 1.
DR PROSITE; PS50844; AFP_LIKE; 1.
SQ SEQUENCE 346 AA; 35052 MW; C167A8A5A0A2721D CRC64;

Query Match 69.7%; Score 1239; DB 2; Length 346;
Best Local Similarity 68.9%; Pred. No. 2.3e-79;
Matches 237; Conservative 48; Mismatches 59; Indels 0; Gaps 0;

Qy 1 MSNIYVAEIGCHNGSVDIAREMILKAKAGVNAVFOETKADKLISAPRAEYOJIN 60
Db 1 MSNIYVAEIGCHNGSVDIAREMILKAKAGVNAVFOETKADKLISAPRAEYOJIN 60
Qy 61 TGELESQLEMTKLEMKYDDYHMEYAVSLNDVSTPDESDIDFLASKOKIMKIPS 120
Db 61 TGELESQLEMTKLEMKYDDYHMEYAVSLNDVSTPDESDIDFLASKOKIMKIPS 120
Qy 121 GELINLPYLEKIAKLPIPDCKIIISTGMATIDEIKOSVSIFINNKVAVGNTITLHCNTY 180
Db 121 GELINLPYLEKIAKLPIPDCKIIISTGMATIDEIKOSVSIFINNKVAVGNTITLHCNTY 180
Qy 181 PTFEPEDVNAINDLKHPKNNIGFSDHSGFYAAIAAPVIGITTEKFTLTKMSGP 240
Db 181 PTFEPEDVNAINDLKHPKNNIGFSDHSGFYAAIAAPVIGITTEKFTLTKMSGP 240
Qy 241 DHIASIEPDELKHLICIGVRCVKSLSGNSKVVTASERKNKIVAKSIIAKTEIKGEVFS 300
Db 241 DHIASIEPDELKHLICIGVRCVKSLSGNSKVVTASERKNKIVAKSIIAKTEIKGEVFS 300
Qy 301 EKNTTRPGNGISPMEMWNLGKIAEODFIDELIHSEFNOE 346
Db 301 EKNTTRPGNGISPMEMWNLGKIAEODFIDELIHSEFNOE 346

Qy 241 DHIASIEPDELKHLICIGVRCVKSLSGNSKVVTASERKNKIVAKSIIAKTEIKGEVFS 300
Db 241 DHIASIEPDELKHLICIGVRCVKSLSGNSKVVTASERKNKIVAKSIIAKTEIKGEVFS 300
Qy 301 EKNTTRPGNGISPMEMWNLGKIAEODFIDELIHSEFNOE 344
Db 301 EKNTTRPGNGISPMEMWNLGKIAEODFIDELIHSEFNOE 344

RESULT 3

Q87186 PRELIMINARY; PRT; 341 AA.
AC Q87186 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE NEUB protein.
GN NEUB.
OS Streptococcus agalactiae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1311;
RN NCBI_TaxID=1311;
RP SEQUENCE FROM N.A.
RA Yamamoto S., Miyake K., Iijima S.;
RT "Identification and Characterization of cps (capsular polysaccharide)
genes from Streptococcus agalactiae Type Ia.";
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Miyake K., Yamamoto S., Koike Y., Watanabe M., Iijima S.;
RT "Molecular Characterization of Type-Specific Capsular Polysaccharide
Biosynthesis Genes of Streptococcus agalactiae Type Ia.";
RL Submitted (JUN-1999) to the EMBL/Genbank/DBJ databases.
DR EMBL; AB017355; BA33753.1; -;
DR EMBL; AB028896; BA48287.1; -;
DR GO; GO:0016051; P:carbohydrate biosynthesis; IEA.
DR InterPro; IPR006014; Antifreeze dom.
DR InterPro; IPR006190; Antifreeze-like.
DR Pfam; PF01354; Antifreeze; 1.
DR Pfam; PF01354; Antifreeze; 1.
DR PROSITE; PS50844; AFP_LIKE; 1.
SQ SEQUENCE 341 AA; 38091 MW; 17620BF82A2BA340 CRC64;

Query Match 55.2%; Score 981.5; DB 2; Length 341;
Best Local Similarity 56.3%; Pred. No. 3.2e-61;
Matches 192; Conservative 55; Mismatches 89; Indels 5; Gaps 2;

Qy 4 IYIYVAEIGCHNGSVDIAREMILKAKAGVNAVFOETKADKLISAPRAEYOJIN 63
Db 4 IYIYVAEIGCHNGSVDIAREMILKAKAGVNAVFOETKADKLISAPRAEYOJIN 63
Qy 64 LESQLEMTKLEMKYDDYHMEYAVSLNDVSTPDESDIDFLASKOKIMKIPS 123
Db 64 LESQLEMTKLEMKYDDYHMEYAVSLNDVSTPDESDIDFLASKOKIMKIPS 123
Qy 122 TNLPLFLKIGK---QOKAVILSTGVAWEEIHOAVNLRQGT--DISILHCTEYPTP 176
Db 122 TNLPLFLKIGK---QOKAVILSTGVAWEEIHOAVNLRQGT--DISILHCTEYPTP 176
Qy 184 PEDVNAINDLKHPKNNIGFSDHSGFYAAIAAPVIGITTEKFTLTKMSGP 243
Db 184 PEDVNAINDLKHPKNNIGFSDHSGFYAAIAAPVIGITTEKFTLTKMSGP 243
Qy 237 ASATPDIADALVKGRIYEGALGREKIPDPVEENKIVAKSIIAKTEIKGEVFS 296
Db 237 ASATPDIADALVKGRIYEGALGREKIPDPVEENKIVAKSIIAKTEIKGEVFS 296
Qy 304 ITTKRPGNGISPMEMWNLGKIAEODFIDELIHSEFNOE 344
Db 304 ITTKRPGNGISPMEMWNLGKIAEODFIDELIHSEFNOE 344

RESULT 4

Q8E501 PRELIMINARY: PRT; 341 AA.
 ID Q8E501
 AC Q8E501
 DT 01-MAR-2003 (TREMBlrel. 23, Created)
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
 DE Hypothetical protein.
 GN NEUB OR GBS1236.
 OS Streptococcus agalactiae (serotype III).
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 CC Streptococcus.
 NCBI_TaxID=216495;
 RX [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MEM316 / Serotype III;
 RA Glaser P., Ruenick C., Buchrieser C., Chevalier F., Frangoul L.,
 RA Masdek T., Zouine M., Couve E., Lalloué L., Poyart C., Tieu-Cuot P.,
 RA Kunst F.,
 RT "Genome sequence of Streptococcus agalactiae, a pathogen causing
 RT invasive neonatal disease.".
 RL Mol. Microbiol. 45:1499-1513(2002).
 DR EMBL; AL766849; CAD46895.1; .
 DR SAGelList; gbs1236; .
 DR GO; GO:0016051; P:carbohydrate biosynthesis; IEA.
 DR InterPro; IPR006014; Antifreeze dom.
 DR InterPro; IPR006190; Antifreeze-like.
 DR InterPro; IPR004144; Neut.
 DR Pfam; PF01354; Antifreeze; 1.
 DR Pfam; PF03102; Neut; 1.
 DR PROSITE; PS50844; AFP_LIKE; 1.
 DR Hypothetical protein; Complete proteome.
 KW SEQUENCE 341 AA; 38091 MM; 176208PF82BA340 CRC64;
 SQ

Query Match 55.1%; Score 981.5; DB 16; Length 341;
 Best Local Similarity 56.3%; Pred. No. 3.2e-61;
 Matches 192; Conservative 55; Mismatches 89; Indels 5; Gaps 2;

QY 4 IYVAEIGCHNGSVDIAREMILKAEAGVNAVKQTFKADKLISALAPKAEYQKNTGE 63
 DB 2 VYIIAEIGCHNGSDINAKKWDVAVSCGVDAVKQTFKAEKLSKPAKAEYQKNTGT 61
 QY 64 LESQLEMTKLEMKYDYDYLHMEYAVSLNDVFTSPDESDIDFLASIKQKIMKIPSGEL 123
 DB 62 ADSQLEMTKRLLESEFEYLEMDYALSKGVETFTSPDEBSLEFLISTDMPIYKIPSGEI 121
 QY 124 LNEPYLEKIAKLPIDPKKIIISTGMATIDEIKOSVSIFFINNKVPVGNITLHCNTEYPTP 183
 DB 122 TNLPLEKIKG--QQKVVILSTGMAVMEIHQAVNILLRONGTT--DISILHCTTEYPTP 176
 QY 124 FEDVNLAINDIKKEFPKNNIGFSDHSSGFYAALAAVPGITFEKFTLDSKSGPDHL 243
 DB 177 YPSLNLAVHITLKDSEFQDLTIGSDHSIGSEVPILAAAGAEVIEKHTLDTNNMGDPDK 236
 QY 244 ASIEPDELKHLCTIGRCVSKSLSGNSKVTASERKNTVAKSIIAKTEIKKGVPSEKN 303
 DB 237 ASATPDIILAAVKGVRIVEQALGREFKIPDPVEEKNTVARKSVVALKPIKKGDIYSTEN 296
 QY 304 ITTKRPNNGISPMWYNILGKLAEDFIPDELIIHSEFKQ 344
 DB 237 ITTKRPNNGISPMWYNILGKLAEDFIPDELIIHSEFKQ 337

RESULT 5
 Q9ALM6 PRELIMINARY: PRT; 339 AA.
 ID Q9ALM6
 AC Q9ALM6
 DT 01-JUN-2003 (TREMBlrel. 17, Created)
 DT 01-JUN-2003 (TREMBlrel. 17, Last sequence update)
 DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)

DE Putative N-acetyl neuraminic acid synthetase Neut.
 GN NEUB.
 OS Streptococcus agalactiae.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 CC Streptococcus.
 NCBI_TaxID=1311;
 RX [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NT6;
 RA McKinnon K., Chaffin D.O., Rubens C.E.;
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF357958; AAK1670.1; .
 DR GO; GO:0016051; P:carbohydrate biosynthesis; IEA.
 DR InterPro; IPR006014; Antifreeze dom.
 DR InterPro; IPR006190; Antifreeze-like.
 DR InterPro; IPR004144; Neut.
 DR Pfam; PF01354; Antifreeze; 1.
 DR Pfam; PF03102; Neut; 1.
 DR PROSITE; PS50844; AFP_LIKE; 1.
 DR PROSITE; PS50844; AFP_LIKE; 1.
 KW SEQUENCE 339 AA; 37773 MM; 156C2DBEPID453C7 CRC64;
 SQ

Query Match 55.1%; Score 979.5; DB 2; Length 339;
 Best Local Similarity 56.3%; Pred. No. 4.4e-61;
 Matches 192; Conservative 54; Mismatches 90; Indels 5; Gaps 2;

QY 4 IYVAEIGCHNGSVDIAREMILKAEAGVNAVKQTFKADKLISALAPKAEYQKNTGE 63
 DB 2 VYIIAEIGCHNGSDINAKKWDVAVSCGVDAVKQTFKAEKLSKPAKAEYQKNTGT 61
 QY 64 LESQLEMTKLEMKYDYDYLHMEYAVSLNDVFTSPDESDIDFLASIKQKIMKIPSGEL 123
 DB 62 ADSQLEMTKRLLESEFEYLEMDYALSKGVETFTSPDEBSLEFLISTDMPIYKIPSGEI 121
 QY 124 LNEPYLEKIAKLPIDPKKIIISTGMATIDEIKOSVSIFFINNKVPVGNITLHCNTEYPTP 183
 DB 122 TNLPLEKIKG--QQKVVILSTGMAVMEIHQAVNILLRONGTT--DISILHCTTEYPTP 176
 QY 184 FEDVNLAINDIKKEFPKNNIGFSDHSSGFYAALAAVPGITFEKFTLDSKSGPDHL 243
 DB 177 YPSLNLAVHITLKDSEFQDLTIGSDHSIGSEVPILAAAGAEVIEKHTLDTNNMGDPDK 236
 QY 244 ASIEPDELKHLCTIGRCVSKSLSGNSKVTASERKNTVAKSIIAKTEIKKGVPSEKN 303
 DB 237 ASATPDIILAAVKGVRIVEQALGREFKIPDPVEEKNTVARKSVVALKPIKKGDIYSTEN 296
 QY 304 ITTKRPNNGISPMWYNILGKLAEDFIPDELIIHSEFKQ 344
 DB 297 ITTKRPNNGISPMWYNILGKLAEDFIPDELIIHSEFKQ 337

RESULT 6
 Q93T12 PRELIMINARY: PRT; 339 AA.
 ID Q93T12
 AC Q93T12
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
 DE Putative N-acetyl neuraminic acid synthetase Neut.
 GN NEUB.
 OS Streptococcus agalactiae.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 CC Streptococcus.
 NCBI_TaxID=1311;
 RX [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CNMC 1/82;
 RA McKinnon K., Chaffin D.O., Rubens C.E.;
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF355776; AAK43615.1; .
 DR GO; GO:0016051; P:carbohydrate biosynthesis; IEA.
 DR InterPro; IPR006014; Antifreeze dom.
 DR InterPro; IPR006190; Antifreeze-like.
 DR InterPro; IPR004144; Neut.

DR
EMU; 1404555; AREA 5001.1; -
PIR; 144651; 144651.

DR InterPro; IPR006190; Antifreeze_1ike.

DR InterPro; IPR006190; Antifreeze_1ike.

DR InterPro: IPR004144; Neut.
 DR Pfam: PF01354; Antifreeze; 1.
 DR Pfam: PF01302; Neut; 1.
 DR PROSITE: PS50844; AFP_LIKE; 1.
 KW Complete proteome.
 SQ SEQUENCE 341 AA; 38033 MW; 1483756C2B2BF314 CRC64;

Query Match 55.18; Score 979.5; DB 16; Length 341;
 Best Local Similarity 56.38; Pred. No. 4.4e-61;
 Matches 192; Conservative 54; Mismatches 90; Indels 5; Gaps 2;

QY 4 IYIVAEICGNHNSVDIAREMILKAKAGVAVKQFETKADKLSALAPKAEYQIKRTGE 63
 DB 2 VYIIAEICGNHNSINLAKKAVDVAVSCGVDAVVFQTFKAEKLSKPAKAEYQKATGT 61
 QY 64 LESQLEMTKLEMKYDYDLHMEYAVSLNDVSTPDEDSIDFLASLKQIKWIPSGEL 123
 DB 62 ADSQLEMTKRLLETFEYLEMRDYALSKGVETFPSTPDEDSLEFLISTDMPYIKPSGEI 121
 QY 124 LNTPLYEKIATLPIDPKKIISTGMATIDEIKOSVSIFINNKVPVGNITLIHCNTEYPTP 183
 DB 122 TMLPYLEKIGK--QOKKVIISTGMAMERITHQAVNIIHQNGTT--DISTLHCTTEYPTP 176
 QY 184 FEDVNLAINDLKHPNNIGFSDHSGFYAALAAVPYGITTFLEKHTLIDKSMGPDHL 243
 DB 177 YPSLNLNVHITLKDPEKDLITIGSDHSGSEVPYIAAAMGAEVIEKFTLDTNMEGPDHK 236
 QY 244 ASIEPDELKHLCTIGRCVCKSLGNSKVYASERKAKIVAKSIIATKTIKKGEVSEK 303
 DB 237 ASATPDILALVGVKIVYEQALGREFEKIPDPEERKAKIVAKSVVALKPKIKGDIYSIEN 296
 QY 304 IITKRPNGISPMENYVNLGKIAEQDFIPDELIHSEFPKQ 344
 DB 297 IITKRPNGISPMENYVNLGKIAEQDFIPDELIHSEFPKQ 337

RESULT 9

Q9A016 PRELIMINARY; PRT; 341 AA.
 ID Q9A016;
 AC Q9A016;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Neut.
 GN Neut.
 OS Streptococcus agalactiae.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 CC Streptococcus.
 NCBI_TaxID=1311;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Miyake K., Matanabe M., Iijima S.;
 RT "GenJ of Streptococcus agalactiae type 1b shows beta-1,3-
 RT galactosyltransferase activity";
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB050723; BAB21605.1;
 DR GO; GO:0016051; P:carbohydrate biosynthesis; IEA.
 DR InterPro; IPR006014; Antifreeze_dom.
 DR InterPro; IPR006190; Antifreeze_like.
 DR InterPro; IPR004144; Neut.
 DR Pfam; PF01354; Antifreeze; 1.
 DR Pfam; PF01302; Neut; 1.
 DR PROSITE; PS50844; AFP_LIKE; 1.
 SQ SEQUENCE 341 AA; 38037 MW; 300AEB4925D2F9D CRC64;

Query Match 53.58; Score 951.5; DB 2; Length 341;
 Best Local Similarity 54.88; Pred. No. 4.2e-59;
 Matches 187; Conservative 55; Mismatches 94; Indels 5; Gaps 2;
 QY 4 IYIVAEICGNHNSVDIAREMILKAKAGVAVKQFETKADKLSALAPKAEYQIKRTGE 63
 DB 2 VYIIAEICGNHNSINLAKKAVDVAVSCGVDAVVFQTFKAEKLSKPAKAEYQKATGT 61

QY 64 LESQLEMTKLEMKYDYDLHMEYAVSLNDVSTPDEDSIDFLASLKQIKWIPSGEL 123
 DB 62 ADSQLEMTKRLLETFEYLEMRDYALSKGVETFPSTPDEDSLEFLISTDMPYIKPSGEI 121
 QY 124 LNTPLYEKIATLPIDPKKIISTGMATIDEIKOSVSIFINNKVPVGNITLIHCNTEYPTP 183
 DB 122 TMLPYLEKIGK--QOKKVIISTGMAMERITHQAVNIIHQNGTT--DISTLHCTTEYPTP 176
 QY 184 FEDVNLAINDLKHPNNIGFSDHSGFYAALAAVPYGITTFLEKHTLIDKSMGPDHL 243
 DB 177 YPSLNLNVHITLKDPEKDLITIGSDHSGSEVPYIAAAMGAEVIEKFTLDTNMEGPDHK 236
 QY 244 ASIEPDELKHLCTIGRCVCKSLGNSKVYASERKAKIVAKSIIATKTIKKGEVSEK 303
 DB 237 ASATPDILALVGVKIVYEQALGREFEKIPDPEERKAKIVAKSVVALKPKIKGDIYSIEN 296
 QY 304 IITKRPNGISPMENYVNLGKIAEQDFIPDELIHSEFPKQ 344
 DB 297 IITKRPNGISPMENYVNLGKIAEQDFIPDELIHSEFPKQ 337

RESULT 10

Q9F508 PRELIMINARY; PRT; 344 AA.
 ID Q9F508;
 AC Q9F508;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Putative N-acetyl neuraminic acid synthetase Neut (EC 4.1.3.-).
 GN NAB1 OR LA1609.
 OS Leptospira interrogans.
 OC Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.
 CC Bacteria; Spirochaetes;
 NCBI_TaxID=173;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN=56601 / Serogroup Icterohaemorrhagiae / Serovar 1a1;
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB011338; AAN48808.1;
 DR GO; GO:0016829; P:lyase activity; IEA.
 DR GO; GO:0016051; P:carbohydrate biosynthesis; IEA.
 DR InterPro; IPR006014; Antifreeze_dom.
 DR InterPro; IPR006190; Antifreeze_like.
 DR InterPro; IPR004144; Neut.
 DR Pfam; PF01354; Antifreeze; 1.
 DR Pfam; PF01302; Neut; 1.
 DR PROSITE; PS50844; AFP_LIKE; 1.
 KW Complete proteome.
 SQ SEQUENCE 344 AA; 38200 MW; 985DFOE0340F1792 CRC64;

Query Match 48.88; Score 868; DB 16; Length 344;
 Best Local Similarity 52.78; Pred. No. 3.2e-53;
 Matches 116; Conservative 54; Mismatches 100; Indels 4; Gaps 2;
 QY 3 NIYIVAEICGNHNSVDIAREMILKAKAGVAVKQFETKADKLSALAPKAEYQIKRTG 62
 DB 13 NIIILIERGVNNEHGDNNILAEKILNVAFDGVADVQTFHATETLANSKPKADYQISMNG 72
 QY 63 ELESQLEMTKLEMKYDYDLHMEYAVSLNDVSTPDEDSIDFLASLKQIKWIPSGE 122
 DB 73 EGGTQVSMKKELESVQDHFRLIEICKKNVIGPLSTAFULKSIDLLIEGLSLMKIPSGE 132
 QY 123 LNTPLYEKIATLPIDPKKIISTGMATIDEIKOSVSIFINNKVPVGNITLIHCNTEYPTP 182
 DB 133 ITNYPYKIKIGEL--NOKIILISTGMANLGEISALNVINSGSSLSNITLIHCNTEYPTP 189
 QY 183 FEDVNLAINDLKHPNNIGFSDHSGFYAALAAVPYGITTFLEKHTLIDKSMGPDHL 242
 DB 190 FFEVNLVAMETLKSAF-PAKVGYSDHTVGIIVAAVAVAGVAVIEKHFTLDTNMEGPDH 248
 QY 243 ASIEPDELKHLCTIGRCVCKSLGNSKVYASERKAKIVAKSIIATKTIKKGEVSEK 302
 DB 249 KASLEBNEKLTWSSIRNTEKMGDGIKPPSSERRNIIARKSIVATRNISKGIIFSES 308

QY 303 NITTKRPGNISPMENYINLIGKIAEDFIIDELI 336
DB 309 NITTKRPGDELSPMNMWYVIGKQARDFIIDELI 342

RESULT 11

Q9PMX2 PRELIMINARY; PRT; 334 AA.
AC Q9PMX2;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE N-acetylneuraminic acid synthetase (EC 4.1.3.-).
GN NEUB3 OR C01327.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Campylobacteraceae; Campylobacter.
OX NCBI_TaxID=197;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NOTC 11168;
RX MEDLINE=20150912; PubMed=10686204;
RA Parkhill J., Wren B.W., Murgall K., Ketley J.M., Churcher C.,
RA Basham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,
RA Jagers K., Kariyasev A.V., Moule S., Pallen M.J., Penn C.W.,
RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
RA Whitehead S., Barrall B.G.;
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
reveals hypervariable sequences.";
RL Nature 403:665-668(2000).
DR EMBL; AL139078; CAB73754.1; -.
DR PIR; D81276; D81276.
DR GO; GO:0004190; P:aspartic-type endopeptidase activity; IEA.
DR GO; GO:0005524; P:ATP binding; IEA.
DR GO; GO:0006508; P:carbohydrate biosynthesis; IEA.
DR GO; GO:0006014; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR006190; Antifreeze_dom.
DR InterPro; IPR001969; Asprotease_AS.
DR InterPro; IPR005479; CPase_L_D2.
DR InterPro; IPR004144; Neub.
DR Pfam; PF03182; Neub; I.1.
DR PROSITE; PS00844; ASP_LIKE; 1.
DR PROSITE; PS00141; ASP_PROTEASE; 1.
DR PROSITE; PS00867; CPASASE_2; 1.
KM Complete proteome.
SQ SEQUENCE 334 AA; 37091 MW; 59FA32E8EA92CFD CRC64;

Query Match 48.8%; Score 867; DB 16; Length 334;
Best Local Similarity 51.5%; Pred. No. 3.7e-53;
Matches 173; Conservative 62; Mismatches 97; Indels 4; Gaps 2;

QY 1 MSNIYVAEIGCHNHSVDIAREMILKAEAGVNAVKQTFKADKLISAIAPEAYQIKN 60
DB 1 MKKTLILAEAGVNHNDLNLAKKLILADSGADPFKQSPAKNISTKAKAPQYLT 60
QY 61 TGLBSQLAMTKLEMKYDDYDLHMEYAVSLNDVSTPPDESDIDFLASLKQKIWKIPS 120
DB 61 TANDESOLOMWOGLIEDLNAHKELILHAKKCNIAFSTPPDESVDILNEGLIKFKIPS 120
QY 121 GELIANTPYLEKIAKLPIPDKKIISTGMATIDELIKOSVIFINNKVPVGNITILHCNTEY 180
DB 121 GELIANTPYLEKIAKLPIPDKKIISTGMATIDELIKOSVIFINNKVPVGNITILHCNTEY 180
QY 121 GELIANTPYLEKIAKLPIPDKKIISTGMATIDELIKOSVIFINNKVPVGNITILHCNTEY 180
DB 121 GELIANTPYLEKIAKLPIPDKKIISTGMATIDELIKOSVIFINNKVPVGNITILHCNTEY 180
QY 181 PTPPEVDNALINDLKKHPPKNNIGFSDHSGFYAALAAVPGYITTEHFTLDSKSGSP 240
DB 181 PTPPEVDNALINDLKKHPPKNNIGFSDHSGFYAALAAVPGYITTEHFTLDSKSGSP 240
QY 178 PAPFNVNKKAMOSLDAF-KLDVGYSDHTGRIHISLAVALGACVIEGHFTLDSKSGSP 236
DB 178 PAPFNVNKKAMOSLDAF-KLDVGYSDHTGRIHISLAVALGACVIEGHFTLDSKSGSP 236
QY 241 DHLASIEPDELKHLGICVRCVSKLSGNSKVYATASERKNKIVARSKIATKEIKKEGVFS 300
DB 241 DHLASIEPDELKHLGICVRCVSKLSGNSKVYATASERKNKIVARSKIATKEIKKEGVFS 300
QY 237 DHLASIEPDELKHLGICVRCVSKLSGNSKVYATASERKNKIVARSKIATKEIKKEGVFS 296
DB 237 DHLASIEPDELKHLGICVRCVSKLSGNSKVYATASERKNKIVARSKIATKEIKKEGVFS 296

QY 301 ENITTKRPGNISPMENYINLIGKIAEDFIIDELI 336
DB 297 ENITTKRPGDELSPMNMWYVIGKQARDFIIDELI 332

RESULT 12

Q7X523 PRELIMINARY; PRT; 334 AA.
AC Q7X523;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE PTMC.
GN PTMC.
OS Campylobacter coli.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Campylobacteraceae; Campylobacter.
OX NCBI_TaxID=195;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC167;
RX MEDLINE=96423180; PubMed=8825781;
RA Query P., Doig P., Alm R.A., Burr D.H., Kinsella N., Trust T.J.;
RT "Identification and characterization of genes required for post-
translational modification of Campylobacter coli VC167 flagellin.";
RL Mol. Microbiol. 19:369-378(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=VC167;
RA Logan S.M., Kelly J.F., Thibault P., Ewing C.P., Query P.;
RT "Structural heterogeneity of carbohydrate modifications affects
serospecificity of Campylobacter flagellins.";
RL Submitted (JUN-2003) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=VC167;
RA Query P.M., Doig P., Alm R.A., Burr D.H., Kinsella N., Trust T.J.;
RT Submitted (MAY-1995) to the EMBL/Genbank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=VC167;
RA Query P., Ewing C.P., Moran A.P., Trust T.J.;
RT Submitted (OCT-1999) to the EMBL/Genbank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=VC167;
RA Logan S.M., Kelly J.F., Thibault P., Ewing C.P., Query P.;
RT Submitted (MAY-2002) to the EMBL/Genbank/DBJ databases.
DR EMBL; AY102621; NAM76227.1; -.
SQ SEQUENCE 334 AA; 36905 MW; C415F8473D91D186 CRC64;

Query Match 47.5%; Score 844; DB 2; Length 334;
Best Local Similarity 49.7%; Pred. No. 1.5e-51;
Matches 167; Conservative 67; Mismatches 98; Indels 4; Gaps 2;

QY 1 MSNIYVAEIGCHNHSVDIAREMILKAEAGVNAVKQTFKADKLISAIAPEAYQIKN 60
DB 1 MKKTLILAEAGVNHNDLNLAKKLILADSGADPFKQSPAKNISTKAKAPQYLT 60
QY 61 TGLBSQLAMTKLEMKYDDYDLHMEYAVSLNDVSTPPDESDIDFLASLKQKIWKIPS 120
DB 61 TANDESOLOMWOGLIEDLNAHKELILHAKKCNIAFSTPPDESVDILNEGLIKFKIPS 120
QY 121 GELIANTPYLEKIAKLPIPDKKIISTGMATIDELIKOSVIFINNKVPVGNITILHCNTEY 180
DB 121 GELIANTPYLEKIAKLPIPDKKIISTGMATIDELIKOSVIFINNKVPVGNITILHCNTEY 180
QY 121 GELIANTPYLEKIAKLPIPDKKIISTGMATIDELIKOSVIFINNKVPVGNITILHCNTEY 180
DB 121 GELIANTPYLEKIAKLPIPDKKIISTGMATIDELIKOSVIFINNKVPVGNITILHCNTEY 180
QY 181 PTPPEVDNALINDLKKHPPKNNIGFSDHSGFYAALAAVPGYITTEHFTLDSKSGSP 240
DB 181 PTPPEVDNALINDLKKHPPKNNIGFSDHSGFYAALAAVPGYITTEHFTLDSKSGSP 240
QY 178 PAPFNVNKKAMOSLDAF-KLDVGYSDHTGRIHISLAVALGACVIEGHFTLDSKSGSP 236
DB 178 PAPFNVNKKAMOSLDAF-KLDVGYSDHTGRIHISLAVALGACVIEGHFTLDSKSGSP 236
QY 241 DHLASIEPDELKHLGICVRCVSKLSGNSKVYATASERKNKIVARSKIATKEIKKEGVFS 300
DB 241 DHLASIEPDELKHLGICVRCVSKLSGNSKVYATASERKNKIVARSKIATKEIKKEGVFS 300

Db 237 DRAKSLBDELQJCTKIRELESALGDCIGKASRKERNINIAKSLVAKKIKKGISIFS 296
 QY 301 EXNITTKRPNGISPMENYVNLGKIAEODPIPBELI 336
 Db 297 ENNLTKRPAGISAMRYDEVLGKKASRDVEDELI 332

RESULT 13

Q70911 PRELIMINARY; PRT; 333 AA.

AC Q70911; PRT; 333 AA.
 DT 01-OCT-2003 (Tremblrel. 25, Created)
 DT 01-OCT-2003 (Tremblrel. 25, Last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
 DT Putative N-acetylneuraminic acid synthetase (BC 4.1.3.-).
 GN NEUB, SPSE OR SYNW0448.
 OS Synchococcus sp. (strain WH8102).
 OC Bacteria; Cyanobacteria; Chroococcales; Synchococcus.
 NC NCB1_Taxid=84588;
 RN [1]
 SEQUENCE FROM N.A.
 RX MEDLINE=22825697; PubMed=12917641;
 RA Patenik B., Brahamsa B., Larimer F.W., Land M., Hauser L., Chain P.,
 RA Lamerdin J., Regala W., Allen E.E., McCarten J., Paulsen I.,
 RA Durene A., Patensky F., Webb E.A., Waterbury J.,
 RT "The genome of a motile marine Synchococcus".
 RL Nature 424:1037-1042 (2003).
 DR EMBL, BX569690; CAB06963.1;
 KW Lyase; Complete proteome.
 SQ SEQUENCE 333 AA; 36219 MW; 2859DC1763C94E CRC64;

Query Match 43.3%; Score 769; DB 16; Length 333;
 Best local Similarity 45.5%; Pred. No. 2,9e-46;
 Matches 153; Conservative 62; Mismatches 117; Indels 4; Gaps 2;

QY 1 MSNITVAEIGCNHNSVDIAEMILKAEKAGVNAVKQTFKADKLISALPAKAYOIK 60
 Db 1 MOKTLVIAEAGVNHNSVDIAEMILKAEKAGVNAVKQTFKADKLISALPAKAYOIK 60
 QY 61 TGELESQLEMTKLEMKYDDYLHMEYAVSLNDVSPFPDESDIDFLASKOKIKIP 120
 Db 61 LDSEEGQALMKLELPEHVELEISCKNIFLSTAFISISGLATLKPKNKVP 120
 QY 121 GELNLPEYLEKIAKLPIPDKKIISTGATIDEIKOSVSIFINNKPVGNIITLAC 180
 Db 121 GEINLPEYLRIGRQGP--VILSTGANLGEIBALSVLEQKGTFRSMITVLACT 177
 QY 181 PTPEDVNLAINDLKHPKPNIGSDHSGFYAALAAVPGITFEKFTLDKMSGP 240
 Db 178 PAPVQVNLKRAMNSISQAFDV-AVGYSDBHSGISVPIAAVAGVLEKLLTAR 236
 QY 241 DTLASIEPDELKHLCTGRCVCKSGNSKVYTAERKNIIVAKSIIATKIKGEVFS 300
 Db 237 DTKASLEPDDQFAAVVQGIIRTEQLGDIKRPISSEGNIPYVRKSLVAACPI 296
 QY 301 EXNITTKRPNGISPMENYVNLGKIAEODPIPBELI 336
 Db 297 ENNLTKRPAGISAMRYDEVLGKKASRDVEDEMI 332

RESULT 14

Q8BDZ7 PRELIMINARY; PRT; 333 AA.

AC Q8BDZ7; PRT; 333 AA.
 DT 01-MAR-2003 (Tremblrel. 23, Created)
 DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
 DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
 DE Stalic acid synthase.
 GN VVI0808.
 OS Vibrio vulnificus.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 VC Vibrionaceae; Vibrrio.

OX NCB1_Taxid=672;
 RN [1]
 SEQUENCE FROM N.A.
 RC STRAIN=CMCP6;
 RA Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,
 RA Choy H.B.;
 RT "Complete genome sequence of Vibrio vulnificus CMCP6".
 RL Submitted (DEC-2002) to the EMBL/GenBank/DBD databases.
 DR EMBL: AB016799; AAC09312.1;
 DR GO: GO:0016051; P:carbohydrate biosynthesis; IEA.
 DR InterPro: IPR006014; Antifreeze_dom.
 DR InterPro: IPR004149; Antifreeze_like.
 DR Pfam: PF01354; Antifreeze_1.
 DR Pfam: PF03102; Neub_1.
 DR PROSITE: PSS0844; AFP_LIKE; 1.
 KW Complete proteome.
 SQ SEQUENCE 333 AA; 36504 MW; 419D1FLA07EB32A4 CRC64;

Query Match 40.7%; Score 723.5; DB 16; Length 333;
 Best local Similarity 44.2%; Pred. No. 4.7e-43;
 Matches 149; Conservative 66; Mismatches 117; Indels 5; Gaps 3;

QY 1 MSNITVAEIGCNHNSVDIAEMILKAEKAGVNAVKQTFKADKLISALPAKAYOIK 59
 Db 1 MTKTLVIAEAGVNHNSVDIAEMILKAEKAGVNAVKQTFKADKLISALPAKAYOIK 60
 QY 60 NTGELESQLEMTKLEMKYDDYLHMEYAVSLNDVSPFPDESDIDFLASKOKIKIP 119
 Db 61 NTGELESQLEMTKLEMKYDDYLHMEYAVSLNDVSPFPDESDIDFLASKOKIKIP 119
 QY 120 SGELNLPEYLEKIAKLPIPDKKIISTGATIDEIKOSVSIFINNKPVGNIITLAC 179
 Db 120 SGELNLPEYLRHNSPA--KPVILSTGMKYSVEHAAVLTADAGSLDMITVLACT 176
 QY 180 YTPEDVNLAINDLKHPKPNIGSDHSGFYAALAAVPGITFEKFTLDKMSGP 239
 Db 177 YTPEDVNLAINDLKHPKPNIGSDHSGFYAALAAVPGITFEKFTLDKMSGP 236
 QY 240 PDHLASIEPDELKHLCTGRCVCKSGNSKVYTAERKNIIVAKSIIATKIKGEVFS 299
 Db 237 PDHLASIEPDELKHLCTGRCVCKSGNSKVYTAERKNIIVAKSIIATKIKGEVFS 296
 QY 300 EXNITTKRPNGISPMENYVNLGKIAEODPIPBELI 336
 Db 297 CADMLEIKRPNGISPTRWDEVGSIKADYOIGELI 333

RESULT 15

Q87770 PRELIMINARY; PRT; 357 AA.

AC Q87770; PRT; 357 AA.
 DT 01-JUN-2003 (Tremblrel. 24, Created)
 DT 01-JUN-2003 (Tremblrel. 24, Last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
 DE Putative N-acetylneuraminic acid synthetase.
 GN VP0290.
 OS Vibrio parahaemolyticus.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 VC Vibrionaceae; Vibrrio.
 NC NCB1_Taxid=670;
 RN [1]
 SEQUENCE FROM N.A.
 RC STRAIN=RIMD 2210633 / Serotype O3:k6;
 RX MEDLINE=22508454; PubMed=12620739;
 RA Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tegomori K.,
 RA Iijima Y., Nishijima M., Nakano M., Yamashita A., Kubota Y., Kimura S.,
 RA Yasunaga T., Honda T., Shingawa H., Hattori M., Iida T.;
 RT "Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism
 distinct from that of V. cholerae".
 RL Lancet 361:743-749 (2003).
 DR EMBL: AP005073; BAC58463.1;
 DR GO: GO:0016051; P:carbohydrate biosynthesis; IEA.

Thu May 13 11:53:14 2004

us-09-930-440b-8.rsp

Page 8

DR InterPro: IPR006014; Antifreeze_dom.
DR InterPro: IPR006190; Antifreeze_like.
DR InterPro: IPR004144; Neut.
DR Pfam: PF01354; Antifreeze; 1.
DR Pfam: PF01102; Neut; 1.
DR PROSITE: PS00844; AFP_like; 1.
KM Complete proteome.
SQ SEQUENCE 357 AA; 38639 MW; 8B887C5568BC8F54 CRC64;

Query Match 39.8%; Score 708.5; DB 16; Length 357;
Best Local Similarity 43.1%; Pred. No. 5.9e-42;
Matches 154; Conservative 60; Mismatches 114; Indels 29; Gaps 4;

QY 6 IVAEICNNGSVDAIREMILKAKAGVNAVKQTEKADKLISAIAPKAEYQKNGELE 65
DB 4 IIEAGVNHNGQELAPALVDNAHEGADIVKQTEKAKNLVTEAKQAEYQVTKQGE 63
QY 66 SQLEMTKLEMKYDYLHMEYAVSINLDVSTPFQEDSIDPLAS-LKQIKWIPSGELL 124
DB 64 SQLAMLSRLSEIHYHDLVKYCNLSIGIEFLSTAFDLESIDPLVNDLGLTRLPSGEEL 123
QY 125 NEPLYEKIATLPFOKKIISTGMATIDEIKOSVIF----- 161
DB 124 NAEPLVAHAARTGC--DLIVSTGMATLSELEALGVIAFGYTAODEKPSMLGFQEAIVAS 180
QY 162 -INNKVPVGNITTLHCTEYPTPEEDVNLNAILDKHPPKNNIGFSDHSSGFYAAIAAV 220
DB 181 EAGQKALKEKVTLLHCTEYPAFMAEINLSMDTLGSAFDL-AAGYSDHSEGITPIAAV 239
QY 221 PYGITTEKFTLDKSNSGPDHLASIHPPDELKHLCLGVRCVEKSGSNKVTATSEKPK 280
DB 240 ARGAVLIEKFTLDKNNBGPDKASLEPQELGAWKAIRQIEVALSSVKTPTESSEYK 299
QY 281 IVAKSIIAKTEIKGKGVFSEKNTTTRPGNGISPMETWNLGKIAEODFIPELII 337
DB 300 AVAKSLVAATDIKSGELTELNLTKRPSGLSPYTWELIGSKATKGKAGELL 356

Search completed: May 6, 2004, 09:07:37
Job time : 36.1171 secs

Thu May 13 11:53:13 2004

us-09-930-440b-8.ra1

Page 1

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OK protein - protein search, using sw model

Run on: May 6, 2004, 08:59:54 / Search time 14.4061 seconds
(without alignments)
1239.929 Million cell updates/sec

Title: US-09-930-440B-8

Perfect score: 1778

Sequence: 1 MSNIYVAIGCNHNGSVDI.....EQDFIPDELIIHSEFKNGE 346

Scoring table:

BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: Issued Patents AA:*

1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCTUS.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	510.5	28.7	359	4	US-09-516-143A-4
2	475.5	26.7	346	4	US-09-495-406-21
3	110.5	6.2	1365	4	US-09-376-330-18
4	109.5	6.2	739	4	US-09-543-681A-6437
5	108.5	6.1	1365	6	5194600-4
6	101	5.7	772	4	US-09-543-681A-6467
7	99.5	5.6	10182	4	US-09-134-001C-3159
8	98.5	5.5	373	4	US-09-724-864-43
9	98	5.5	256	4	US-09-107-532A-4072
10	96.5	5.4	525	3	US-09-113-750A-35
11	94	5.3	491	4	US-09-914-259-18
12	93	5.2	815	4	US-09-543-681A-6286
13	93	5.2	1155	4	US-08-480-604A-28
14	92.5	5.2	1296	1	US-08-405-496A-28
15	92.5	5.2	1296	3	US-08-915-136-28
16	92.5	5.2	1296	4	US-09-084-517-28
17	92.5	5.2	1296	3	US-08-961-083-106
18	92	5.2	626	4	US-09-536-784-106
19	92	5.2	769	4	US-09-789-657-15
20	92	5.2	770	2	US-08-209-521-15
21	92	5.2	770	3	US-08-961-810-123
22	92	5.2	770	3	US-08-352-902D-123
23	92	5.2	770	4	US-09-265-503B-123
24	92	5.2	770	4	US-08-809-326A-6
25	91.5	5.1	571	3	US-09-689-914A-6
26	91.5	5.1	571	4	US-09-689-913A-6
27	91.5	5.1	571	4	US-09-689-913A-6

28	91.5	5.1	571	4	US-09-689-916A-6	Sequence 6, Appli
29	91.5	5.1	580	4	US-09-198-452A-332	Sequence 332, App
30	91	5.1	461	4	US-09-345-473E-35	Sequence 35, Appl
31	90.5	5.1	760	4	US-09-107-532A-5490	Sequence 5490, Ap
32	90	5.1	458	4	US-08-158-452A-209	Sequence 209, App
33	89.5	5.0	693	3	US-08-235-636C-72	Sequence 72, App
34	89	5.0	566	4	US-09-134-001C-3431	Sequence 3431, Ap
35	89	5.0	580	4	US-08-913-159-10	Sequence 10, Appl
36	89	5.0	751	4	US-09-252-991A-33073	Sequence 33073, A
37	89	5.0	1087	4	US-09-914-259-12	Sequence 12, Appl
38	88.5	5.0	854	4	US-09-254-352B-18	Sequence 18, Appl
39	88.5	5.0	992	4	US-09-206-942-61	Sequence 59, Appl
40	88.5	5.0	998	4	US-09-206-942-59	Sequence 59, Appl
41	87.5	4.9	563	4	US-09-107-532A-4741	Sequence 4741, Ap
42	87	4.9	392	1	US-07-768-286B-4	Sequence 4, Appli
43	87	4.9	406	4	US-09-252-991A-22251	Sequence 22251, A
44	87	4.9	477	2	US-08-432-016-3	Sequence 3, Appli
45	87	4.9	477	2	US-08-684-594-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-09-516-143A-4
Sequence 4, Application US/09516143A
Patent No. 633182
GENERAL INFORMATION:
APPLICANT: Human Genome Sciences, Inc.
TITLE OF INVENTION: Human Glycosylation Enzymes
FILE REFERENCE: P505PCT
CURRENT APPLICATION NUMBER: US/09/516, 143A
CURRENT FILING DATE: 2000-03-01
PRIOR APPLICATION NUMBER: 60/122,409
PRIOR FILING DATE: 1999-03-02
NUMBER OF SEQ ID NOS: 6
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 4
LENGTH: 359
TYPE: PRT
ORGANISM: Homo sapiens
US-09-516-143A-4

Query Match	28.7%	Score 510.5;	DB 4;	Length 359;
Best Local Similarity	36.3%	Pred. No. 2.3e+46;		
Matches	123;	Conservative	61;	Mismatches 140; Indels 15; Gaps 6;
QY	5	YIVAIGCNHNGSVDIAREMILKAKERGVNAKQTRFADKLISAIAPKAEYQIKNGEL	64	
DB	20	FLIAEIQNHQGDIDVAKRMIRMAKCEQADCAKQSKELBFKFKALKERPYTSKHSWG	78	
QY	65	ESQLEMTKKLEMKYDDYHLKRYAVSLIDVSTPDESDIDPLASTLKOKIWKIPSGEL	124	
DB	79	KTYGHRKHLFSDQIRELQRYAEVGIIFTLASGDEMAYELHENVPPFKVSGSDTN	138	
QY	125	NLPYLEKTIAPLEPKKIIISTGATIDEIKOSVIFINNTVPVGNITLHONTBYPPE	184	
DB	139	NFPYLEKTIAPLEPKKIIISTGATIDEIKOSVIFINNTVPVGNITLHONTBYPPE	192	
QY	185	EDVNNALINDKPKPKKNNIGFSDHSGSFVAIAAVPGIFIEKFTLDSKMSGPDHIA	244	
DB	193	EDVNNALINDKPKPKKNNIGFSDHSGSFVAIAAVPGIFIEKFTLDSKMSGPDHIA	252	
QY	245	SIEPDELKHLGICVACVKSJGSNSKVYTAEROKKIIVAKRSIIAKTBIRKGEVFSKNT	304	
DB	253	SIEPDELKHLGICVACVKSJGSNSKVYTAEROKKIIVAKRSIIAKTBIRKGEVFSKNT	312	
QY	305	TTK-RPENGISPKEMVYLTK-----IAEDFIPDELI	336	
DB	313	TVKVGEP-KAYPPEDEIPLVAKKVLVTVEEDDTIWERLV	350	

RESULT 2

Thu May 13 11:53:13 2004

us-09-930-440b-8.rat

Page 2

```
US-09-495-406-21
; Sequence 21, Application US/09495406
; Patent No. 6503744
; GENERAL INFORMATION:
; APPLICANT: Gilbert, Michel
; APPLICANT: Makarchuk, Warren W
; APPLICANT: National Research Council of Canada
; TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
; TITLE OF INVENTION: Gangliosides and Ganglioside Mimics
; FILE REFERENCE: 019633-000110US
; CURRENT APPLICATION NUMBER: US/09/495,406
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: US 60/118,213
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: Patent Ver. 2.1
; SEQ ID NO 21
; LENGTH: 346
; TYPE: PRT
; ORGANISM: Campylobacter jejuni
; FEATURE:
; OTHER INFORMATION: sialic acid synthase from C. jejuni OH4384 (ORF 8a
US-09-495-406-21
Query Match
Best Local Similarity 26.7%; Score 475.5; DB 4; Length 346;
Best Local Similarity 36.7%; Pred. No. 1,3e-42;
Matches 124; Conservative 54; Mismatches 147; Indels 13; Gaps 7;
QY 6 IVAEIGCHNNGSVIAREMILKAEAGVNAVKPQTFKADGLISALPKAEYQIKNTGEL 65
DB 19 VPEFIGNHNSLEIAKIMVDAFSTGAKIKQTHIVEDEMSKAKV---IPGNAKI- 74
QY 66 SOLMTKLEMKYDDYVLMMEYVNSLNDVSTPPEBDSIDPLASLKQKIMTPSGEELN 125
DB 75 STYEMOCALDYKDELAKETKELGLVYLSLTPFSAGANRLKEDMGAFAFISGECNN 134
QY 126 LPYLEKIALPIPDKKIISTGMATIDIKOSVIFINNKVFNVTITLHCNTEYPTPE 185
DB 135 YPLKHLIAF---KKPMIVSTRGNISISIKPTVKILNDIEIP---FVLMHTNLTVPPTPN 188
QY 186 DYNLAIHNDLKGFPPKNIGSGDSGFAIAAVPYGTFIEKHPITLIDKSNQPDHLAS 245
DB 189 LVNLAMLEKKER-SCWVGLSDHTTNLALGAVLAGCVLERHTDSMHSRSGEDVCS 247
QY 246 IEPDELKHCIGRCVCKSLGSN-SKVTYASERKNTVAKSIIATKELKGEVPSKNT 304
DB 248 MOTQALKEIITQSSQKALMGNNESKRAAKQEQVTIDFAFASVYSIKDIKGEVLSMNI 307
QY 305 TTKRFG-NGISPMEMYNLGLKIAEQDFIPDELIHSEF 341
DB 308 WKRPGLGIGIAAEFENILGKKALRDIENDTQSYEDF 345
RESULT 3
US-09-376-330-18
; Sequence 18, Application US/09376330
; Patent No. 6393321
; GENERAL INFORMATION:
; APPLICANT: Tessier, Daniel C.
; APPLICANT: Dispart, Daniel
; APPLICANT: Bergeron, John J.M.
; APPLICANT: Thomas, David Y.
; TITLE OF INVENTION: Method for screening for
; TITLE OF INVENTION: UDP-glucose:glycoprotein glucosyltransferase (UGGT) activity
; TITLE OF INVENTION: and nucleic acid encoding for UGGT
; FILE REFERENCE: 2139-9US
; CURRENT APPLICATION NUMBER: US/09/376,330
; PRIOR FILING DATE: 1999-08-18
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 18
; LENGTH: 1365
```

```
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: S. cerevisiae KRES
US-09-376-330-18
Query Match
Best Local Similarity 6.2%; Score 110.5; DB 4; Length 1365;
Best Local Similarity 21.5%; Pred. No. 0.03; Indels 97; Gaps 19;
Matches 78; Conservative 60; Mismatches 127; Indels 97; Gaps 19;
QY 62 GELESQLEMTK--KLEMKYDYVLMMEYVNSLNDVSTPPEBDSIDPLASLKQKIMTP 119
DB 41 GVDNQL-LTNLYPVLGDLDDIDQENLVALNTVLRERYDEKADVADLLELYASLY--P 97
QY 120 SGEL-----LNLPLEKIALPIPD-----KKIISTGMATIDE 153
DB 98 MAMIGDISNAEQDANSYFVLANKRYK-----PDVFLKSDLTLOQKVPDQV 151
QY 154 IK-QSVSIFINNKVFNVTITLHCNTEYPTPEEDVNLN---AIN-DLKHPKKNIGS 207
DB 152 IEPYDVIGTNSKADI--LILYGPVIDSDFEFENLPMEMNGKFRFTMRSTCSL 209
QY 208 DSSGKY-----AIAAVPYGTFIEK-HFTLDKSM---SGPDHLASIEPD 249
DB 210 DKSVEYPLTHLEITLNGSRMSIPQ---LKKILYTPKELVAGANDDQDLHLEPE 265
QY 250 ELKHLICIGRCVCKSLGSNSKVYVNSERKIVAR-----KSIATKELKGEVSEK 303
DB 266 ELRELDLRTSLISSEFYQKDTATLMTFISVNNPPLISKQIKVSVNKOITNSNE 325
QY 304 ITRK-----RPNGISPMEM-----YNLQ-----KIAS-QDFIDELIHS 339
DB 326 LNSKGEVNLGLVYNGQNKITSLETPNLTALKTEYQSLKLTMLQELPSKCLDS 385
QY 340 EF 341
DB 386 KP 387
RESULT 4
US-09-543-681A-6437
; Sequence 6437, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BREON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRA
; FILE REFERENCE: 2709,1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; PRIOR FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 6437
; LENGTH: 739
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-6437
Query Match
Best Local Similarity 6.2%; Score 109.5; DB 4; Length 739;
Best Local Similarity 19.1%; Pred. No. 0.014;
Matches 70; Conservative 56; Mismatches 127; Indels 113; Gaps 14;
QY 37 KQTRKADLISAIAPK-AYQIKNTGELSGOLEMT-----KLEMKY 78
DB 102 KMTTSLNDEBNTKPSLAERPKKKODHDVSAQNSLPTSSPDLKKKKNKY 161
QY 79 DDTLHMAVYNSLNDVSTPPEBDSIDPLASLKQKIMTPSGELNPLTEKIALP- 136
DB 162 DFLPLQSLQVHIDILINKIKDEKEDY-----YRDLQQLSNG 201
QY 137 IPDKKIISTGMATIDRIK-----QSVSIFINN-KVP----- 167
```


Db 202 IDDLKELFKSGFHELDTKAPVADYVMEYVNOIKENKPKSEYKSTLNNCNIPDSLIQQ 261
 QY 168 --VGNITLHCNTE-----YPTFEDVNNALNDLKKHPKNNIGSPDSHSG 212
 Db 262 LILGNT---HKKVMKALBDEVPFLNFPSPGSLSDLDLNDLSKLD---GPFDSFN 314
 QY 213 FYAIAAAPPYGTPIETKHPFLDKMSGPDHLASIEPD---ELKHLICIGVRCVSKLG 266
 Db 315 YLPISISEDKINKITSHTELEKALINKKESSEIKQLABIRGLFKELITISQSSLEDEIA 374
 QY 267 SNS---KVTASERKKKIYAKSIIAKTEIKKEGVSEKNIITKRRGNGISPEWENL 322
 Db 375 SNSISGMSLTFPLSKRELTLLKVMHRSESSQ-KLHNQGVTEK-----SLK 422
 QY 323 GKIAEQ 328
 Db 423 DKVEEQ 428
 RESULT 5
 5194600-4
 Patent No. 5194600
 APPLICANT: BUSSEY, HOWARD; BOONE, CHARLES; SOMMER, STEVE S.;
 HILL, KATHRYN; MEADEN, PHILIP
 TITLE OF INVENTION: GENES WHICH PARTICIPATE IN B-GLUCAN
 ASSEMBLY AND USE THEREOF
 NUMBER OF SEQUENCES: 4
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/488, 316
 FILING DATE: 05-MAR-1990
 SEQ ID NO: 4
 LENGTH: 1365
 5194600-4

Query Match 6.1%; Score 108.5; DB 6; Length 1365;
 Best Local Similarity 21.3%; Pred. No. 0.049;
 Matches 77; Conservative 61; Mismatches 127; Indels 97; Gaps 19;
 QY 62 GELSEOLEMTK--KLEMKYDDYLLHMEYAVSINLDVSTPDEDSIDFLSKKIKIP 119
 Db 41 GDVNNQL-LTNLYPLVTGLDEIDIDIOENTVALTSNVRERYKEDVDADLELYASLY-P 97
 QY 120 SGSL-----LNLPLYEKIAPLPD-----KKIISTGQATIDE 153
 Db 98 MCMQHDDISNAEDDANSSYFVLANGNRYEK-----PDVFLKSKDLTIQKVPVDV 151
 QY 154 IK-QSVSIFINNKVPVGNITTLHCNTEYPTFEDVNNALNDL---KHPKNNIGS 207
 Db 152 IQPYDVVIGTNSBAPI--LILYGCPTVIDSDPEEFNRRLFHEANNGEGKFRFWRSTCSL 209
 QY 208 DHSSGFY-----AIAAVPGYGTPIETK-HETLDKS--MSGPDH---LASIEPD 249
 Db 210 DKQVEYELTHPLBTLTONGSRMSIFQ---LKKILYVKEKILVADNKKQJHDELE 265
 QY 250 ELKHLICIGVRCVSKLGSNSKVTASERKKKIYAR---KSLIAETIKKEGVSEKNI 303
 Db 266 ELRELKLVSTLISEFYQKDIATLNTFKSIYVNEPLSKKOLIKVSVAKDITITNEE 325
 QY 304 ITTK-----RPGNGISPEWENL-----YNLG-----KTAE--ODPIPELIIHS 339
 Db 326 LNSGQFDVNMGLYINGQNMKITSLTPYNNLTALKTEYQSLKITTNLQLEPESKCLIDS 385
 QY 340 EF 341
 Db 386 KF 387
 RESULT 6
 US-09-543-681A-6467
 Sequence 6467, Application US/09543681A
 Patent No. 6605709
 GENERAL INFORMATION:
 APPLICANT: GARY BRETON

QY 79 DDYLLHMEYAVSINLDVSTPDEDSIDFLSKKIKIPSGEL-----LNLPLYEK 131
 Db 70 EDYIAPGDY--SLSVYNTPEPLGEKTIILPKANKEGV--VPVTTIDKLKGVNVALSK 125
 QY 132 IANLPDPPDKKIISTGQATIDEIKQSVSIFINNKVPVGNITTLHCNTEYPTFEDVNNL 150
 Db 126 LSDLP-EDTEL-----DNIEAVIPSRSLTFNINKLTLNASF--QVMMDK 167
 QY 191 -----AINDLKHPKNNI--GPDHSGGFYAAI-AAVPGYGTPIEK 230
 Db 168 SFQGEIPAEILLDGVSAVILMYANFTKRTDYSSHNKNNFLNAGVNIQPMWRLTN 227
 QY 231 FTLDKMSGPDHLASIEPDLEKHLICIGVRCVSKLGSNSKVTASERKKKIYAKSIIAX 290
 Db 228 YSYNNS--SGSDQKSSSD-----SESNYVVRNIAVLYK 259
 QY 291 TEIKKEGVSEKNIITKRRGNGIS 314
 Db 260 SMLKIGHITGGNIFSSIPKIGIS 283
 RESULT 7
 US-09-134-001C-3159
 Sequence 3159, Application US/09134001C
 Patent No. 6380370
 GENERAL INFORMATION:
 APPLICANT: Lynn Doucette-Stamm et al
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
 TITLE OF INVENTION: EPIDERMIDS FOR DIAGNOSTICS AND THERAPEUTICS
 FILE REFERENCE: GTC-007
 CURRENT APPLICATION NUMBER: US/09/134, 001C
 CURRENT FILING DATE: 1998-08-13
 PRIOR APPLICATION NUMBER: US 60/064,964
 PRIOR FILING DATE: 1997-11-08
 PRIOR APPLICATION NUMBER: US 60/055,779
 PRIOR FILING DATE: 1997-08-14
 NUMBER OF SEQ ID NOS: 5674
 SEQ ID NO 3159
 LENGTH: 10182
 TYPE: PRT
 ORGANISM: Staphylococcus epidermidis
 US-09-134-001C-3159

Query Match 5.6%; Score 99.5; DB 4; Length 10182;
 Best Local Similarity 17.8%; Pred. No. 13; Indels 87; Gaps 14;
 Matches 59; Conservative 71; Mismatches 87; Gaps 14;
 QY 17 SYDIAEKILAKEA--GVNAVRFQTFKADKLISAI--APKAY-----QIKNGTELESQ 67
 Db 8897 AIDTFKODIKTKDANGIRLTPAKSKAKELDLSKFIKAKQFTHANDEIMATNSI--- 8953
 QY 68 LEKTKLEMKYD-----DYLLHMEYAVSINLDVSTPDEDSIDFLSKKIKIP 115
 Db 8954 AOLSRIVNCAFDLNDAMKSLHDELINQAFPVQASNYTNS--DEDLKQOPDHALSNARKV 9011

QY 116 WKIPSGELLNLPYLEKIAKLPIDPKKIIISTGKATIDELIKOSVIFINNKNVPGVNTIIH 175
 DB 9012 LAKENKNIDKDEKIOGL-KQVIEDTKDAN-GIQRISKAKAKAIQYVQ----- 9057
 QY 176 CATEYTPFEDVNLAINDLKHPFKNIGSPDHSGFYAALA----- 218
 DB 9058 -----SLSTINDKQRIANNTIHSDDLSLANTSKASDLDNNKDLADTIE 9105
 QY 219 ----AVPGITPI-EKHFITD-----KMSGPDLASIEPDEIKHLCTGVRCVE 262
 DB 9106 SNSTSVNSVNYINADKNQIEFEDALQOASATSSKTSSENPATIE--EYAGLSQAIYDK 9163
 QY 263 KSGSNSKVTASERKOKIYARSKIATKEIK 294
 DB 9164 NALNGORLATERKSKOLKIKGLKDLNKQALD 9195

RESULT 8
 US-09-724-864-43
 Sequence 43, Application US/09724864
 Patent No. 6380362
 GENERAL INFORMATION:
 APPLICANT: Watson, James D.
 APPLICANT: Marison, James G.
 TITLE OF INVENTION: Polynucleotides, polypeptides expressed
 TITLE OF INVENTION: by the polynucleotides and methods for their use.
 FILE REFERENCE: 11000,105001
 CURRENT APPLICATION NUMBER: US/09/724,864
 CURRENT FILING DATE: 2000-11-28
 PRIOR APPLICATION NUMBER: U.S. No. 6380362 60/171,678
 PRIOR FILING DATE: 1999-12-23
 NUMBER OF SEQ ID NOS: 72
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 43
 LENGTH: 373
 TYPE: PRT
 ORGANISM: Mouse
 US-09-724-864-43

Query Match 5.5%; Score 98.5; DB 4; Length 373;
 Best Local Similarity 21.2%; Pred. No. 0.069;
 Matches 63; Conservative 48; Mismatches 83; Indels 103; Gaps 16;
 QY 68 LEMTKLEKMYDYHLME-YAVSLNLDVFSPPEDSDIDPLASLKOKIWKIPSGEL--- 123
 DB 102 LONNKIKEIKENDFKSITSLIALILNNKLTIKHPK---TFLTKRLRLYLSHNQLSEI 158
 QY 124 -LNLPLEKIAKLPIDPKKIIISTGKATIDELIKOSVIFINNKNVPGVNTIIH 174
 DB 159 PLVLP--KSLAEKRIHDKMYKRIQKDTFGSMALHVEWSANPLENNGIEPAGFEGVTF 216
 QY 175 HCNTEYTPFEDVNLAINDLKHPFKNIGSPDHSGFYAALAIVPGIT--TFLKHF 232
 DB 217 HIRIA-----EAKTISIPKGLPPTLEHLD 242
 QY 233 LBSMSGPHLASIEPDEIKHLCTGVRCVE-SLGSNSKVTASERKOKIYARSKIAT 291
 DB 243 FNR-----ISTVELDLR---YRELQRLGLGNN-----RIT 271
 QY 292 EIKG-----EYFSEKNITTKRPGNGISPMEW-----YNLKGIAEQDPIF 332
 DB 272 DINGTFANIPRREHLHBNKIKLP-SGLDELKYLQIIFLHNSIAKGVNDPFCP 327

RESULT 9
 US-09-107-532A-4072
 Sequence 4072, Application US/09107532A
 Patent No. 6583275
 GENERAL INFORMATION:
 APPLICANT: Lynn A Doucette-Stamm and David Bush
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
 ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
 NUMBER OF SEQUENCES: 7310

CORRESPONDENCE ADDRESS:
 ADDRESSEE: GENOME THERAPEUTICS CORPORATION
 STREET: 100 Beaver Street
 CITY: Waltham
 STATE: Massachusetts
 COUNTRY: USA
 ZIP: 02154
 COMPUTER READABLE FORM:
 MEDIUM TYPE: CD-ROM ISO9660
 COMPUTER: PC
 OPERATING SYSTEM: <Unknown>
 SOFTWARE: ASCII
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/107,532A
 FILING DATE: 30-Jun-1998
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 60/085,598
 FILING DATE: 14 May 1998
 APPLICATION NUMBER: 60/051571
 FILING DATE: July 2, 1997
 ATTORNEY/AGENT INFORMATION:
 NAME: Arinello, Pamela Deneke
 REGISTRATION NUMBER: 40,489
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (781)893-5007
 TELEFAX: (781)893-8277
 INFORMATION FOR SEQ ID NO: 4072:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 256 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHEICAL: YES
 ORIGINAL SOURCE:
 ORGANISM: Enterococcus faecium

FEATURE:
 NAME/KEY: misc_feature
 LOCATION: (B) LOCATION 1..256
 SEQUENCE DESCRIPTION: SEQ ID NO: 4072:
 US-09-107-532A-4072

Query Match 5.5%; Score 98; DB 4; Length 256;
 Best Local Similarity 25.7%; Pred. No. 0.042;
 Matches 47; Conservative 26; Mismatches 76; Indels 34; Gaps 7;
 QY 29 KEAGNAVAFQTFKADKLISAIAPAEYQIKNTGELBSQI-EMTKLEKMYDYHLMEY 87
 DB 101 KQGVNKAKFVYD---VSDILQMDVYKDLTTQIESQLDRYKLS---SDQLKLVSN 153
 QY 88 AVSLNLDVFSPPEDSDIDPLASLKOKIWKIPSGELNLPYLEKIAKLPIDPKKIIIST 146
 DB 154 GETKITIDQGVF-EFSVEVQNDQAVYIVNESSPEITSNSEDILLITPVSQDN----- 207
 QY 147 GKATIDELIKOSVIFINNKNVPGVNTIIHCNTEYTPFEDVNLAINDLKHPFKNIGF 206
 DB 208 -----GEPKQVWYI-----PSSMSQPEBVEYKIVSTGVKNFFANCMDF 248
 QY 207 SDH 209
 DB 249 VTH 251

RESULT 10
 US-09-113-750A-35
 Sequence 35, Application US/09113750A
 Patent No. 6294176
 GENERAL INFORMATION:
 APPLICANT: David E. Junker and Mark D. Cochran
 TITLE OF INVENTION: Recombinant Raccoonpox virus
 NUMBER OF SEQUENCES: 58
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: John P. White

Thu May 13 11:53:13 2004

us-09-930-440b-8.ra1

Page 5

STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/113,750A
FILING DATE:
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 55744
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)262-0400
TELEFAX: (212)664-0525
TELEX: 422523
INFORMATION FOR SEQ ID NO: 35:
SEQUENCE CHARACTERISTICS:
LENGTH: 525 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-113-750A-35

Query Match 5.4%; Score 96.5; DB 3; Length 525;
Best Local Similarity 23.5%; Pred. No. 0.2;
Matches 50; Conservative 36; Mismatches 96; Indels 31; Gaps 8;

QY 12 CHNHSVDAIEEMILKAEAGVNAVKFQTFRAKLIISAIAPKAYQIINNGELSEOLEMT 71
DB 172 CHHSYVFKGECIDNIYEDNNTKIEPNF--DSLSN--LTDAYVIRVTDASTQIWEK 227
QY 72 KKLKMKYDYIHLMEYANSLADVFSTPDESDIDFLASLKQKTKIPSGELNL---- 126
DB 228 KSVLDRYVESYSHRYSKSHYFKGFSYVRKDDIM--NIVKEP--LSNGASLTINDGSR 283
QY 127 --PYLEKIAKLPIDPKIIISTGMATIDEIKOSVSIFINN-----KVPYGNITIL 174
DB 284 WDPILIVRRRTIIMLMDIINNHTTIDKCIYIHIYINNYRNDYPPFRCLVITN---K 340
QY 175 HCNTEYPTPEFEDV--NINAINDLKHPKONI 204
DB 341 HCLAKYTNHDDIYGTPLHMLSNKGLITPYVI 373

RESULT 11
US-09-314-268-3
Sequence 3, Application US/09314268
Patent No. 6346377
GENERAL INFORMATION:
APPLICANT: Doobar, John
TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO SCREENING FOR PAPILLOMA
TITLE OF INVENTION: VIRUSES
FILE REFERENCE: 3789/80902
CURRENT APPLICATION NUMBER: US/09/314,268
CURRENT FILING DATE: 1999-03-19
EARLIER APPLICATION NUMBER: 09/314,268
EARLIER FILING DATE: 1999-05-18
NUMBER OF SEQ ID NOS: 179
SOFTWARE: Patent in Ver. 2.1
SEQ ID NO 3
LENGTH: 491
TYPE: PRT
ORGANISM: Homo sapiens
US-09-314-268-3

Query Match 5.3%; Score 94; DB 4; Length 491;
Best Local Similarity 23.2%; Pred. No. 0.33;
Matches 78; Conservative 38; Mismatches 128; Indels 92; Gaps 14;

QY 49 AIAPKAEYQIKNGELSEOLEMTKLEMKYDYLHMEYAVSLNDVYST-----P 99
DB 135 AIAKPHIITATPERLIDHLENTKGRNLAKTLVDEARILNMD--FETEVKILKIVP 193
QY 100 FDESDIDFLASLKQKTKIPSGELNLPYLEKIAKLPIDPKIIISTGMATIDEIKOSVS 159
DB 194 RDRKTFEFSALMTKVKQKGRALKN-----PV---KCAVSSKYQTEVDAQ--YY 239
QY 160 IFYNNKVP-----VGNITLHCNTEYPTPEFEDVNLAINDLKHPKNNIGESD 208
DB 240 IFIPSKEDYLYIINELAGNSFMTFCST-----CNYQTALLIRNLGP-- 285
QY 209 HSGGFYAIAAVPYGIFPIKHEFTLDKSNGPDLASIEPDELKHLCT-----GVR 259
DB 286 -----TALPLHQMSK-----RLGSINKFKAKASIIATDVASRGLD 323
QY 260 CEEKSLGNSKVVYASERKKKIYAR-----KSIATKEIKGEVFS--EKNTTKRP 309
DB 324 IPRDYVYVNFIDIPHSKDYIHRVGRTPRAGRSKALITVQYDV--ELFORIEHLIGKLP 382
QY 310 GNGISPEWYNLQKIAE--ODFIDELIHSEPKNO 344
DB 383 GPFQDDDEVNMLTERVAEQRFAPMRLREHGKKKK 418

RESULT 12
US-09-914-259-18
Sequence 18, Application US/09914259
Patent No. 6495336
GENERAL INFORMATION:
APPLICANT: Makowski, Lee
APPLICANT: Hyman, Paul
APPLICANT: Williams, Mark
TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES
FILE REFERENCE: 8471-010-999
CURRENT APPLICATION NUMBER: US/09/914,259
CURRENT FILING DATE: 2000-11-21
NUMBER OF SEQ ID NOS: 180
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 18
LENGTH: 815
TYPE: PRT
ORGANISM: Caenorhabditis elegans
US-09-914-259-18

Query Match 5.2%; Score 93; DB 4; Length 815;
Best Local Similarity 19.9%; Pred. No. 0.98;
Matches 65; Conservative 49; Mismatches 90; Indels 122; Gaps 14;

QY 83 HLMFYAVSLN-----LDVSTPFD--EDSIDFLASLKQKTKIPSGELNLPYLEK 131
DB 96 HTEGVIGDGLSGIIPRIYADIFNHTYSMDENLQF--HKKVSYTEL-----YNEK 144
QY 132 IAKLPIDPKIIISTGMATIDEIKOSVSIFINNKPVGNGITLHCNTEYPTPEFEDVNLAIN 191
DB 145 IRDLDPDEKYNL-----SIHEK-----NRVP-----YKGAITERVGGEDVLOA 185
QY 192 INDLKHPKNNIGFSDHSGFYAIAAVPYGIFPIKHEFTLDKSNGPDLASIER-- 246
DB 186 IEDGKSNRVAAYTNMMEHSRSHSV-----FLITVQEQHTTKQKLGKYLVDLAGSEK 240
QY 249 -----DELKHL-----CIGV-----RCVEKSLGNSK-- 270
DB 241 VSKTGAQGYTLERAKNINKSLTALGIVISALAGTSHPYRDSKLTIRIQESLGGNSRT 300
QY 271 --VVTASER-----KNTIVAKSITAKTEIKGAVFSERK----- 303
DB 301 TVIICASPSHFNBAETKSTLLFGARAKTIQVAVQINDELTAEBMKGRYREKEKENTRIIAA 360

Thu May 13 11:53:13 2004

us-09-930-440b-8.raf

Page 6

QY 304 -----ITTKRPGNGISPMWYVL 321
Db 361 LLOAALELSRMPAGSVSEVENVL 386

RESULT 13
US-09-543-681A-6286
Sequence 6286, Application US/09543681A
Patent No. 6605709
GENERAL INFORMATION:
APPLICANT: GARY BRETTON
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
FILE REFERENCE: 2709, 1002-001
CURRENT APPLICATION NUMBER: US/09/543, 681A
PRIOR FILING DATE: 2000-04-05
PRIOR APPLICATION NUMBER: US 60/128,706
PRIOR FILING DATE: 1999-04-09
NUMBER OF SEQ ID NOS: 8344
SEQ ID NO 6286
LENGTH: 1155
TYPE: PRT
ORGANISM: Proteus mirabilis
US-09-543-681A-6286

Query Match 5.2%; Score 93; DB 4; Length 1155;
Best Local Similarity 20.3%; Pred. No. 1.7;
Matches 69; Conservative 57; Mismatches 110; Indels 104; Gaps 17;

QY 62 GELSSQ--LEMTKLEMKDYHLMYAVSLNLFVSTPPEDSI-DPLSLKQIK 117
Db 85 GDLGQALLDNIQQLKADY-----NKTVEQLPEKRSIQNQLSKQKIAN 134
QY 118 IPGGLMLPYLEKIALPIPDKIIISTGATIDEIKOSV----- 158
Db 135 KEVD-----YQSLESLP-----LATLESQLEVLQSLAKACEDLANYSNELIVLQ 182
QY 159 -----SIFNNKVPVNTI-----LHCTEYTPREDVNAINDLKHF 201
Db 183 QPERAGSVLFNNSEKQIRIALNKKSSADKQSSSVQLLQYLLQSSSFQRTIQ 242
QY 202 NKIGFS-----DHSSGFYALAAVYGITFEKFTL-----DKMSGPDHLA-SIE 247
Db 243 SNVQLSLQLQGRDYSAY-----IDLSQEHQQLQELISKRIDSSEVAKKQ 292
QY 248 PDLKHCICVRC-----VEKSLGNSKVTASERKAKIYAKKSIIAKTEIKGEVS 300
Db 293 TAEISQAIKQNAFYLAQADINKKL--SDKLITTONNEL--NRHSLMYGNRLDRA-109 348
QY 301 EKNTTKRPGNGISPMWYVLGKI--ABODPIPELIIH 338
Db 349 ERHAKQ-----IDVLKGLLSRLFEEOIEHPDPIIFN 383

RESULT 14
US-08-480-604A-28
Sequence 28, Application US/08480604A
Patent No. 5736139
GENERAL INFORMATION:
APPLICANT: KINK, JOHN A.
APPLICANT: THALLEY, BRUCE S.
APPLICANT: PADHYE, NISHA V.
APPLICANT: FIRCA, JOSEPH R.
APPLICANT: STAFFORD, DOUGLAS C.
TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND PREVENTION OF C. DIFFICILE DISEASE
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL, LLP
STREET: 220 MONTGOMERY STREET, SUITE 2200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: UNITED STATES OF AMERICA

QY 265 LG-----SNKVVYASE-----RQKIVA 283
Db 972 SKWKVSLNVEGIIWLODQBEIKQVVERYSOMINISDIYINRMFTVTTNNELNNSKIYI 1031
QY 284 KSIYAKTEIKK-GEVFEKNITTRPG-----NGISPMWYVLKQIAODPIPELIIH 338

ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/480,604A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/422,711
FILING DATE: 14-APR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/405,496
FILING DATE: 16-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,154
FILING DATE: 25-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: INGOLIA, DIANE E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPND-01763
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 1296 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-480-604A-28

Query Match 5.2%; Score 92.5; DB 1; Length 1296;
Best Local Similarity 17.8%; Pred. No. 2.4;
Matches 76; Conservative 71; Mismatches 151; Indels 129; Gaps 17;

QY 5 YVVAIGCNHNSVDIAEMKLIKAYAGVNAKQFTPADKLISALAPAEYQIKNTGEL 64
Db 712 YVIVMLAKVNTQIDILIRK-----KQKEA-----L 736
QY 65 ESQLEMTK-LEMKDYDYLHMEYAVSLNLFVSTPPEDSIDPLASIKQIKIPSEGL 123
Db 737 ENQAEATKALIVYQNTQTEBEKNININIDLSKINESINKMININKPLNQSAYL 796
QY 124 LN-LPY-LEKIAKLPIPDKII-----STGATIDEIKQVSIFNNKVP----- 167
Db 797 NMSMIFGVKRLIEDPDAKLDALIKYIVDNRTGLIGVDRLDKVNNITLSDIPQLSKY 856
QY 168 VGNITILHCTEYFPFEDVNL-----NAINDKKHPKNNIG----- 205
Db 857 VDNQSLSTFEYINNIINTSLNRYNSNHLIDSKRYASKINISKVNFPDIDKQYQL 916
QY 206 FSDHSSGFYALA-AVPGITFEIKHFTLDKSMGPDHLASIEPELRLCIGVCEVS 264
Db 917 FVLESSEKLEVLKNAIVNSYV--ENFSTFWIRIPKFNIS--SLNNEYTIIINCMEVN 971
QY 265 LG-----SNKVVYASE-----RQKIVA 283
Db 972 SKWKVSLNVEGIIWLODQBEIKQVVERYSOMINISDIYINRMFTVTTNNELNNSKIYI 1031
QY 284 KSIYAKTEIKK-GEVFEKNITTRPG-----NGISPMWYVLKQIAODPIPELIIH 338

Thu May 13 11:53:13 2004

us-09-930-440b-8.rat

Page 7

Db 1032 NGRLLDQKPSNLGNHASNINMFKDGCGRDTHRYIWKYFNLFDKELNEKEIKD---LY 1088
QY 339 SEFKNOG 345
Db 1089 DNQNSNG 1095

RESULT 15

US-08-405-496A-28
Sequence 28, Application US/08405496A
Patent No. 5919665

GENERAL INFORMATION:

APPLICANT: WILLIAMS, JAMES A.

TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM

TITLE OF INVENTION: NEUROTOXIN

NUMBER OF SEQUENCES: 30

CORRESPONDENCE ADDRESS:

ADDRESSEE: MEDLEN & CARROLL, LLP

STREET: 220 MONTGOMERY STREET, SUITE 2200

CITY: SAN FRANCISCO

STATE: CALIFORNIA

COUNTRY: USA

ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: FLOPPY disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA: US/08/405,496A

APPLICATION NUMBER: 424

FILING DATE: 16-MAR-1995

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/329,154

FILING DATE: 25-OCT-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/985,321

FILING DATE: 04-DEC-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/429,791

FILING DATE: 31-OCT-1989

ATTORNEY/AGENT INFORMATION:

NAME: INGOLIA, DIANE E.

REGISTRATION NUMBER: 40,027

REFERENCE/DOCKET NUMBER: OPND-01308

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 705-8410

TELEFAX: (415) 397-8338

INFORMATION FOR SEQ ID NO: 28:

SEQUENCE CHARACTERISTICS:

LENGTH: 1296 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-405-496A-28

Query Match 5.2%; Score 92.5; DB 2; Length 1296;

Best Local Similarity 17.8%; Pred. No. 2.4;

Matches 76; Conservative 71; Mismatches 151; Indels 129; Gaps 17;

QY 5 YIVAEIGCHNGSVPIAEEMLIKAEAGVNAVKPOTFRADKLISAIAPKAEYQIKNGEL 64

Db 712 YIVNWLAKVTDILIRK---KMKEA-----L 736

QY 65 ESQLEMTKK-LEMKYDGYLHMEVAVSINLVFSTPDESDIDFLASIKOKIWKIPSGEL 123

Db 737 ENQAEATKALINYNQTEBEKNININIDLSKLNESINKAMININKFLNQCVSYL 796

QY 124 LN-LPY-LEKAKIPDPKKIIT---STGMATIDEIKOSVSIPIANKVP----- 167

Db 1032 NGRLLDQKPSNLGNHASNINMFKDGCGRDTHRYIWKYFNLFDKELNEKEIKD---LY 1088

QY 339 SEFKNOG 345

Db 1089 DNQNSNG 1095

RESULT 15

US-08-405-496A-28

Db 797 MNSMIPYGVARLEDFDASLKDALKYIYDNRGTLGQVRLKXKNVNTLSTDIPQLSKY 856

QY 168 VGNITLHGNTEYPTPEDEVN-----NAINDKKHFPKXNIG----- 205

Db 857 VDNQRLSTFTTEYIKNIINTSLNAYESNHLDJSRYASKNIGSKVNFDPIDKQIOL 916

QY 206 FSDHSGFYAALA-AVPGITPTIEKFTLDKSNAGPDHLASIEPDELKHLCTGVRCEKRS 264

Db 917 FNESSKIVILKNAIVNSMY--ENFSTSPWIRIPKFNST---SLNNEYTLINCENNN 971

QY 265 LG-----SNSKYVTASE-----RKKKIVA 283

Db 972 SGKRVSLNGBELIWTQDQEIQRVVFYKQMINISDYINRWI FVLTNNRLNNSKITYI 1031

QY 284 RKSIIAKTEIKK-GVESEKNITTKRPG---NGISPMWYNLLGKIAEQDFIPDELIH 338

Db 1032 NGRLLDQKPSNLGNHASNINMFKDGCGRDTHRYIWKYFNLFDKELNEKEIKD---LY 1088

QY 339 SEFKNOG 345

Db 1089 DNQNSNG 1095

Search completed: May 6, 2004, 09:08:40

Job time: 16.4061 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:07:45 ; Search time 37.4054 Seconds
(without alignments)
2567.492 Million cell updates/sec

Title: US-09-930-440b-8

Perfect score: 1778
Sequence: 1 MSNIYVAEIGCHNNGSVDI.....EODFIDELIHSEFNNQGE 346

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1140673 seqs, 27566755 residues

Total number of hits satisfying chosen parameters: 1140673

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed.
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1778	100.0	346	US-09-930-440b-8	Sequence 8, Appli
2	993.5	55.9	338	US-09-767-041-24	Sequence 24, Appli
3	516.5	29.0	359	US-10-264-237-2597	Sequence 2597, Ap
4	510.5	28.7	359	US-09-984-205-4	Sequence 4, Appli
5	510.5	28.7	359	US-09-930-440b-8	Sequence 6, Appli
6	475.5	26.7	346	US-09-816-028A-35	Sequence 35, Appli
7	475.5	26.7	346	US-10-303-161-35	Sequence 35, Appli
8	475.5	26.7	346	US-10-303-118-35	Sequence 35, Appli
9	475.5	26.7	346	US-10-303-128-35	Sequence 35, Appli
10	475.5	26.7	346	US-10-303-134-35	Sequence 35, Appli
11	475.5	26.7	346	US-10-303-162-35	Sequence 35, Appli
12	426	24.0	311	US-10-156-761-10910	Sequence 10910, A
13	402	22.6	340	US-09-882-227-490	Sequence 490, App
14	381	21.4	340	US-10-335-977-7625	Sequence 7625, Ap
15	373	21.0	314	US-10-335-977-7625	Sequence 7625, Ap

16	174.5	9.8	123	14	US-10-106-628-6926	Sequence 6926, Ap
17	114	6.4	500	12	US-10-282-122A-53768	Sequence 53768, A
18	112	6.3	335	15	US-10-369-493-8945	Sequence 8945, Ap
19	109	6.1	601	12	US-10-282-122A-55143	Sequence 55143, A
20	107	6.0	1577	15	US-10-369-493-6924	Sequence 6924, Ap
21	107	6.0	1577	15	US-10-369-493-6925	Sequence 6925, Ap
22	107	6.0	1577	15	US-10-369-493-6926	Sequence 6926, Ap
23	103	5.8	653	9	US-09-759-010-2	Sequence 2, Appli
24	103	5.8	732	12	US-10-160-748-4	Sequence 4, Appli
25	103	5.8	950	12	US-10-282-122A-7285	Sequence 47285, A
26	102.5	5.8	567	15	US-10-369-493-1053	Sequence 1053, Ap
27	102	5.7	410	12	US-10-425-114-43092	Sequence 43092, A
28	102	5.7	481	14	US-10-032-565-7372	Sequence 7372, Ap
29	102	5.7	1241	12	US-10-282-122A-51957	Sequence 51957, A
30	101.5	5.7	650	12	US-10-282-122A-74461	Sequence 74461, A
31	101.5	5.7	841	9	US-09-861-451A-30	Sequence 30, Appli
32	101	5.7	455	10	US-09-909-567B-48	Sequence 48, Appli
33	101	5.7	455	15	US-10-094-749-1646	Sequence 1646, Ap
34	101	5.7	468	12	US-10-282-122A-76865	Sequence 76865, A
35	100.5	5.7	645	12	US-10-425-114-38654	Sequence 38654, A
36	100.5	5.7	645	12	US-10-425-114-38654	Sequence 22517, A
37	100.5	5.7	707	12	US-10-424-599-225717	Sequence 21877, A
38	100	5.6	878	15	US-10-369-493-21877	Sequence 52209, A
39	99	5.6	266	12	US-10-282-122A-52209	Sequence 76807, A
40	99	5.6	381	12	US-10-282-122A-76807	Sequence 60665, A
41	99	5.6	866	12	US-10-282-122A-60665	Sequence 7841, Ap
42	99	5.6	962	14	US-10-032-585-7841	Sequence 7841, Ap
43	98.5	5.5	358	14	US-10-319-130-22	Sequence 22, Appli
44	98.5	5.5	373	10	US-09-866-050A-680	Sequence 680, App
45	98.5	5.5	373	14	US-10-319-130-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1
US-09-930-440b-8
Sequence 8, Application US/09930440B
Patient No. US20020142386A1
GENERAL INFORMATION:
APPLICANT: Beedaugh et al.
TITLE OF INVENTION: Engineering Intracellular Sialylation Pathways
FILE REFERENCE: PFS09P2
CURRENT APPLICATION NUMBER: US/09/930,440B
PRIOR FILING DATE: 2001-08-16
CURRENT FILING DATE: 2001-08-16
PRIOR APPLICATION NUMBER: 60/227,579
PRIOR FILING DATE: 2000-08-25
PRIOR APPLICATION NUMBER: 09/516,793
PRIOR FILING DATE: 2000-03-01
PRIOR APPLICATION NUMBER: 60/169,624
PRIOR FILING DATE: 1999-12-08
PRIOR APPLICATION NUMBER: 60/122,582
PRIOR FILING DATE: 1999-03-02
NUMBER OF SEQ ID NOS: 8
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 8
LENGTH: 346
TYPE: PRT
ORGANISM: Escherichia coli
US-09-930-440b-8

Query Match 100.0%; Score 1778; DB 9; Length 346;
Best Local Similarity 100.0%; Pred. No. 4, 1e-161; Indels 0; Gaps 0;
Matches 346; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSNIYVAEIGCHNNGSVDIAREMILKAEAGVAAVVEQTFKADKLSAIPAAVQIKN 60
DB 1 MSNIYVAEIGCHNNGSVDIAREMILKAEAGVAAVVEQTFKADKLSAIPAAVQIKN 60
QY TGBIESOLETKYKLEMTYDYLIHMEYAVSINDVETPEPDESIDPLASIKQWIPIS 120
DB TGBIESOLETKYKLEMTYDYLIHMEYAVSINDVETPEPDESIDPLASIKQWIPIS 120

QY 121 GELNLPYLEKIAKLPIDPKKIIISTGNAITIDEIKOSVIFINNKVPVGNITILHONTY 180
 DB 121 GELNLPYLEKIAKLPIDPKKIIISTGNAITIDEIKOSVIFINNKVPVGNITILHONTY 180
 QY 181 PTFPEVDNAINDLKKEPPKNNIGSDHSSGFYAAIAVPGITFIEKHFTLDSKMSGP 240
 DB 181 PTFPEVDNAINDLKKEPPKNNIGSDHSSGFYAAIAVPGITFIEKHFTLDSKMSGP 240
 QY 241 DFLASIEPDELKHLCTGVCEKSLGNSKVTYASERKKIYAKRSIIAKTEIKGEVFS 300
 DB 241 DFLASIEPDELKHLCTGVCEKSLGNSKVTYASERKKIYAKRSIIAKTEIKGEVFS 300
 QY 301 EKNTTKRPNGISPEWENLGLKAEOPFIDELIHSSEFNQSG 346
 DB 301 EKNTTKRPNGISPEWENLGLKAEOPFIDELIHSSEFNQSG 346

RESULT 2

US-09-767-041-24
 ; Sequence 24, Application US/09767041
 ; Patent No. US20020055168A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Smith, Hilda
 ; TITLE OF INVENTION: STREPTOCOCCUS SUIIS VACCINES AND DIAGNOSTIC TESTS
 ; FILE REFERENCE: 2183-4726
 ; CURRENT APPLICATION NUMBER: US/09/767,041
 ; PRIOR FILING DATE: 2001-01-22
 ; PRIOR APPLICATION NUMBER: PCT/NL99/00460
 ; PRIOR FILING DATE: 1999-07-19
 ; PRIOR APPLICATION NUMBER: EP98202465.5
 ; PRIOR FILING DATE: 1998-07-22
 ; PRIOR APPLICATION NUMBER: EP98202467.1
 ; PRIOR FILING DATE: 1998-07-22
 ; NUMBER OF SEQ ID NOS: 53
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 24
 ; LENGTH: 338
 ; TYPE: PRT
 ; ORGANISM: Streptococcus suis
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; OTHER INFORMATION: CPSP2P
 ; US-09-767-041-24

Query Match 55.9%; Score 993.5; DB 9; Length 338;
 Best Local Similarity 58.4%; Pred. No. 3.5e-86;
 Matches 199; Conservative 45; Mismatches 92; Indels 5; Gaps 2;

QY 4 IYVAEIGCHNNGSVDIAREMILKAEAGVNAVPTQTFRADKLISAIAPKAEYQIKNQGE 63
 DB 2 VYIIAETGCHNNDVHLARQVAVAVDCGVDAKFOERKADLLISKYAPAEYQKTTGSE 61
 QY 64 LESQLEMTKLEMKYDDYLHMEYAVSLNDVSTPDEDSIDPLASLKQKWKIPSGEL 123
 DB 62 SDGQLEMTKLEMKYDDYLHMEYAVSLNDVSTPDEDSIDPLASLKQKWKIPSGEL 121
 QY 124 LNPYLEKIAKLPIDPKKIIISTGNAITIDEIKOSVIFINNKVPVGNITILHONTY 183
 DB 122 TNPYLEKIGR---QAQKVLSTGMAVMDIHOAVKLQNGT--DISTLHNTTYPTP 176
 QY 184 FEDVNAINDLKKEPPKNNIGSDHSSGFYAAIAVPGITFIEKHFTLDSKMSGPDL 243
 DB 177 YPPLNINLVHMLTKKEFPNLTIGYSDHSVGSEVLAIAAAGAELEIKHFTLDSKMSGP 236
 QY 244 ASIEPDELKHLCTGVCEKSLGNSKVTYASERKKIYAKRSIIAKTEIKGEVFS 303
 DB 237 ASATPDIALVGVGRIVEQSLGFEKEPEVEVRNKKIYAKRSIIAKTEIKGEVFS 296
 QY 304 IYKRPNGISPEWENLGLKAEOPFIDELIHSSEFNQSG 344
 DB 297 IYKRPNGISPEWENLGLKAEOPFIDELIHSSEFNQSG 337

RESULT 3

US-10-264-237-2597
 ; Sequence 2597, Application US/10264237
 ; Publication No. US2004009491A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Biese et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 ; FILE REFERENCE: P131p1
 ; CURRENT APPLICATION NUMBER: US/10/264,237
 ; PRIOR FILING DATE: 2002-10-04
 ; PRIOR APPLICATION NUMBER: PCT/US01/16450
 ; PRIOR FILING DATE: 2001-05-18
 ; PRIOR APPLICATION NUMBER: US 60/205,515
 ; PRIOR FILING DATE: 2000-05-19
 ; NUMBER OF SEQ ID NOS: 2876
 ; SOFTWARE: PatentIn Ver. 3.1
 ; SEQ ID NO 2597
 ; LENGTH: 359
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; US-10-264-237-2597

Query Match 29.0%; Score 516.5; DB 15; Length 359;
 Best Local Similarity 36.6%; Pred. No. 1.4e-40;
 Matches 124; Conservative 61; Mismatches 139; Indels 15; Gaps 6;

QY 5 YVAEIGCHNNGSVDIAREMILKAEAGVNAVPTQTFRADKLISAIAPKAEYQIKNQGE 64
 DB 20 FLIAEIGNHGDLDVAKRMIRMAKCGADCAKQKSELERFKALEREYTSKSWG- 78
 QY 65 ESQLEMTKLEMKYDDYLHMEYAVSLNDVSTPDEDSIDPLASLKQKWKIPSGEL 124
 DB 79 KTYGEKHLERSHQYELQRYAEVGIPTASGMDKAEVZLHELNVPEFFKSGSDTN 138
 QY 125 MLPYLEKIAKLPIDPKKIIISTGNAITIDEIKOSVIFINNKVPVGNITILHONTY 184
 DB 139 NFPYLEKIAKGRF---NVISGMSQMDTKQVQIV---KPLNPFCLQCTSAVPLQ 192
 QY 185 EDVNAINDLKKEPPKNNIGSDHSSGFYAAIAVPGITFIEKHFTLDSKMSGPDL 244
 DB 193 EDVNLRAVSEYQKFPDIPIGSGHETGIAISVAVALCAKYLEHITLDTKWSGDISA 252
 QY 245 SIEPDELKHLCTGVCEKSLGNSKVTYASERKKIYAKRSIIAKTEIKGEVFS 304
 DB 253 SIEPDELKHLCTGVCEKSLGNSKVTYASERKKIYAKRSIIAKTEIKGEVFS 312
 QY 305 TTK--RPNGISPEWENLGLKAEOPFIDELIHSSEFNQSG 346
 DB 313 TVKVGEP-KGYPEPDIENLVGRKVLVTVSEDDTIMELV 350

RESULT 4

US-09-984-205-4
 ; Sequence 4, Application US/09984205
 ; Patent No. US20020137175A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Coleman, Timothy A. et al.
 ; TITLE OF INVENTION: Human Glycosyltransferases
 ; FILE REFERENCE: PFS0501
 ; CURRENT APPLICATION NUMBER: US/09/984,205
 ; PRIOR FILING DATE: 2001-10-29
 ; PRIOR APPLICATION NUMBER: PCT/US00/05325
 ; PRIOR FILING DATE: 2000-03-01
 ; PRIOR APPLICATION NUMBER: 09/516,143
 ; PRIOR FILING DATE: 2000-03-01
 ; PRIOR APPLICATION NUMBER: 60/122,409
 ; PRIOR FILING DATE: 1999-03-02
 ; NUMBER OF SEQ ID NOS: 6
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 4
 ; LENGTH: 359
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens

Thu May 13 11:53:13 2004

us-09-930-440b-8.rapb

Page 3

US-09-984-205-4

Query Match 28.7%; Score 510.5; DB 9; Length 359;
Best Local Similarity 36.3%; Pred. No. 5.4e-40;
Matches 123; Conservative 61; Mismatches 140; Indels 15; Gaps 6;

QY 5 YVAAIGCNHNSVDIAREMILAKAGVNAVVFQTFKADKLISALAPKAYQIKNTGEL 64
DB 20 FIIAIGQNHQGBLDVAKRMIRVAKCGADCAFOKSELEFKENKALRPPYTSKSG- 78
QY 65 ESOLMTKLEKMYDYDLHMEYASLNDVFPSTPDESDIDFLASIKQIKWIPSGELL 124
DB 79 KITGEKHKHLEFSDQYRELQRYAEVGIPTFASGMDMAVELHBLNVPFVSGSDTN 138
QY 125 NLPYLEKIAKLPIPDKKIISTGMATIDELIKOSVSIFINNKVFNQNTILHCTEYPTPP 184
DB 139 NFPYLEKTAKGRP---MVISGQSWDTWKQYQIV---KPLNPFCEFLQCTSAVPLDP 192
QY 185 EDVNLAINDLKHPKNNIGFSDHSGFYAALAAVPYGITFEKHTLIDKMSGPDHLA 244
DB 193 EDVNLAVISEYQKLPDIPIGYSGHETGLAISVAVALAKVLERHTLIDKTKSGDSHA 252
QY 245 SIEPDELKHLICIGVRCVKSIGNSKVVYASERKNKIIVAKSIITAKTEIKGEVFSKNI 304
DB 253 SIEPDELALVRSVRLVERALGSPTKQLPCMAKNEKLGKSVVAKVPIEGTILTMDEL 312
QY 305 TTK--RPGNGISPMENYNLGK----IAQDFIDELI 336
DB 313 TVKVGEP--KAYPPEDIFNLVGRKVLVTEEDDTIMEELV 350

RESULT 5

US-09-930-440B-6
Sequence 6, Application US/09930440B
Patent No. US20020142386A1
GENERAL INFORMATION:
APPLICANT: Betenbaugh et al.
TITLE OF INVENTION: Engineering Intracellular Sialylation Pathways
FILE REFERENCE: pfs0992
CURRENT APPLICATION NUMBER: US/09/930.440B
CURRENT FILING DATE: 2001-08-16
PRIOR APPLICATION NUMBER: 60/227,579
PRIOR FILING DATE: 2000-08-25
PRIOR APPLICATION NUMBER: 09/516,793
PRIOR FILING DATE: 2000-03-01
PRIOR APPLICATION NUMBER: 60/169,624
PRIOR FILING DATE: 1999-12-08
PRIOR APPLICATION NUMBER: 60/122,582
PRIOR FILING DATE: 1999-03-02
NUMBER OF SEQ ID NOS: 8
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 6
LENGTH: 359
TYPE: PRT
ORGANISM: Homo sapiens
US-09-930-440B-6

Query Match 28.7%; Score 510.5; DB 9; Length 359;
Best Local Similarity 36.3%; Pred. No. 5.4e-40;
Matches 123; Conservative 61; Mismatches 140; Indels 15; Gaps 6;

QY 5 YVAAIGCNHNSVDIAREMILAKAGVNAVVFQTFKADKLISALAPKAYQIKNTGEL 64
DB 20 FIIAIGQNHQGBLDVAKRMIRVAKCGADCAFOKSELEFKENKALRPPYTSKSG- 78
QY 65 ESOLMTKLEKMYDYDLHMEYASLNDVFPSTPDESDIDFLASIKQIKWIPSGELL 124
DB 79 KITGEKHKHLEFSDQYRELQRYAEVGIPTFASGMDMAVELHBLNVPFVSGSDTN 138
QY 125 NLPYLEKIAKLPIPDKKIISTGMATIDELIKOSVSIFINNKVFNQNTILHCTEYPTPP 184
DB 139 NFPYLEKTAKGRP---MVISGQSWDTWKQYQIV---KPLNPFCEFLQCTSAVPLDP 192

QY 185 EDVNLAINDLKHPKNNIGFSDHSGFYAALAAVPYGITFEKHTLIDKMSGPDHLA 244
DB 193 EDVNLAVISEYQKLPDIPIGYSGHETGLAISVAVALAKVLERHTLIDKTKSGDSHA 252
QY 245 SIEPDELKHLICIGVRCVKSIGNSKVVYASERKNKIIVAKSIITAKTEIKGEVFSKNI 304
DB 253 SIEPDELALVRSVRLVERALGSPTKQLPCMAKNEKLGKSVVAKVPIEGTILTMDEL 312
QY 305 TTK--RPGNGISPMENYNLGK----IAQDFIDELI 336
DB 313 TVKVGEP--KAYPPEDIFNLVGRKVLVTEEDDTIMEELV 350

RESULT 6

US-09-816-028A-35
Sequence 35, Application US/09816028A
Patent No. US20020042369A1
GENERAL INFORMATION:
APPLICANT: Gilbert, Michel
APPLICANT: Makarchuk, Warren W.
TITLE OF INVENTION: National Research Council of Canada
TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
FILE REFERENCE: 019633-000111US
CURRENT APPLICATION NUMBER: US/09/816.028A
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: US 60/118,213
PRIOR FILING DATE: 1999-02-01
PRIOR APPLICATION NUMBER: US 09/495,406
PRIOR FILING DATE: 2000-01-31
NUMBER OF SEQ ID NOS: 49
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 35
LENGTH: 346
TYPE: PRT
ORGANISM: Campylobacter jejuni
FEATURE:
OTHER INFORMATION: sialic acid synthase from C. jejuni OH4384 (ORF 84)
OTHER INFORMATION: of 11pooligosaccharide (LOS) biosynthesis locus)
US-09-816-028A-35

Query Match 26.7%; Score 475.5; DB 9; Length 346;
Best Local Similarity 36.7%; Pred. No. 1.1e-36;
Matches 124; Conservative 54; Mismatches 147; Indels 13; Gaps 7;

QY 6 IYAAIGCNHNSVDIAREMILAKAGVNAVVFQTFKADKLISALAPKAYQIKNTGEL 65
DB 19 VPEIGINNSGLBLAKIVDAFASTGAKTIHQTHIVEDEMSKAAKV---IPGNKI- 74
QY 66 SOLENTKLEKMYDYDLHMEYASLNDVFPSTPDESDIDFLASIKQIKWIPSGELL 125
DB 75 SIEYEMOKCADDYQELAKETKGLVYLTSPSRAGANLEBDGSAFKIGSECN 134
QY 126 LPYLEKIAKLPIPDKKIISTGMATIDELIKOSVSIFINNKVFNQNTILHCTEYPTPP 185
DB 135 YPLIKHIAF---KPEMIVSIGNSISLIPYKILDNELP---FVLHNTNLVYTPPN 188
QY 186 DVNLAINDLKHPKNNIGFSDHSGFYAALAAVPYGITFEKHTLIDKMSGPDHLA 245
DB 189 LVRLNLMALKEF--SCWGLSDHTDNLACGAVALGACVLERHTDSMRSGPDIVS 247
QY 246 IEPDELKHLICIGVRCVKSIGNSKVVYASERKNKIIVAKSIITAKTEIKGEVFSKNI 304
DB 248 KOTQALKEILIOSEQAMIRGNNESSKAKQCOYITDFAFASVSIKDIKGEVLSMDNI 307
QY 305 TTKRPG--NGISPMENYNLGKIAQDFIDELIHSSE 341
DB 308 WYRPELGISIAAEFENILGKALRDIENDTOLSYEDF 345

RESULT 7

US-10-303-161-35
Sequence 35, Application US/10303161

Publication NO. US20030148459A1
 GENERAL INFORMATION:
 APPLICANT: Gilbert, Michel
 APPLICANT: Wakarchuk, Warren W.
 APPLICANT: National Research Council of Canada
 TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
 TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
 FILE REFERENCE: 019633-00011US
 CURRENT APPLICATION NUMBER: US/10/303,161
 PRIOR FILING DATE: 2002-11-21
 PRIOR APPLICATION NUMBER: US/09/816,028
 PRIOR FILING DATE: 2001-03-21
 PRIOR APPLICATION NUMBER: US 60/118,213
 PRIOR FILING DATE: 1999-02-01
 PRIOR APPLICATION NUMBER: US 09/495,406
 PRIOR FILING DATE: 2000-01-31
 NUMBER OF SEQ ID NOS: 49
 SOFTWARE: PatentIn Ver. 2.1
 SEQ ID NO 35
 LENGTH: 346
 TYPE: PRF
 ORGANISM: Campylobacter jejuni
 FEATURE:
 OTHER INFORMATION: stalic acid synthase from C. jejuni OH4384 (ORF 8a
 OTHER INFORMATION: of lipooligosaccharide (LOS) biosynthesis locus)
 US-10-303-161-35

Query Match 26.7%; Score 475.5; DB 14; Length 346;
 Best Local Similarity 36.7%; Pred. No. 1,1e-36;
 Matches 124; Conservative 54; Mismatches 147; Indels 13; Gaps 7;

QY 6 IVAEICNGSVVDIAREMILKAKAGVAVKQTFKADKLISALAPKAEYQIKTGELE 65
 DB 19 VVEEIGNHGSLBLAKIMVDAFSTGAKIKHQTHIVEDMSKAKV---IPGNAKI- 74
 QY 66 SOLEMTKLEMKYDYLHMEYAVSLNLDVSTPFDESDIDFLASIKOKIMKIPSGELN 125
 DB 75 SIYEIMOKCALDYDELAKYTERKGLVYSTPFSRAGARLDEMGSAFKIGSGECNN 134
 QY 126 LPYLEKIAKLPIPDKKIISTGMATIDEIKOSVSIFINNKVPVGNITTLACNTEPTPE 185
 DB 135 YPLIKHIAF---KKPMIVSTGMNISIEIKPTVKILDNELP---FVLMHTTNLYPTPHN 188
 QY 186 DVNLAINDLKKEPPKNNIGFSDHSGFYALAAVPGYITFLKHTTLDKSMGPHLAS 245
 DB 189 LVRLNMLELKEF--SCWGLSDHTTDNLACGVALGACVLERHFTDSMRSGPDVCS 247
 QY 246 IEPDELKHLCTIGRCVCEKSLGNS-SKVVTASERKNKIYAKSITIAKTEIKGEVPEKKNI 304
 DB 248 MDQALKEKLTIOSEGMALMRGNESKKAQOEVTIDFAFASVVSIKDIKGEVLSMDNI 307
 QY 305 TTKRPG-NGISPMWYNLIGKIAEODFIPDELIIHSEF 341
 DB 308 WKRPGLGIGISAEPENILGKALRDIENDTQLSYEDF 345

RESULT 8
 US-10-303-118-35
 Sequence 35, Application US/10303118
 Publication No. US20030157655A1
 GENERAL INFORMATION:
 APPLICANT: Gilbert, Michel
 APPLICANT: Wakarchuk, Warren W.
 APPLICANT: National Research Council of Canada
 TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
 TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
 FILE REFERENCE: 019633-00011US
 CURRENT APPLICATION NUMBER: US/10/303,118
 PRIOR FILING DATE: 2002-11-21
 PRIOR APPLICATION NUMBER: US/09/816,028
 PRIOR FILING DATE: 2001-03-21
 PRIOR APPLICATION NUMBER: US 60/118,213
 PRIOR FILING DATE: 1999-02-01

PRIOR APPLICATION NUMBER: US 09/495,406
 PRIOR FILING DATE: 2000-01-31
 NUMBER OF SEQ ID NOS: 49
 SOFTWARE: PatentIn Ver. 2.1
 SEQ ID NO 35
 LENGTH: 346
 TYPE: PRF
 ORGANISM: Campylobacter jejuni
 FEATURE:
 OTHER INFORMATION: stalic acid synthase from C. jejuni OH4384 (ORF 8a
 OTHER INFORMATION: of lipooligosaccharide (LOS) biosynthesis locus)
 US-10-303-118-35

Query Match 26.7%; Score 475.5; DB 14; Length 346;
 Best Local Similarity 36.7%; Pred. No. 1,1e-36;
 Matches 124; Conservative 54; Mismatches 147; Indels 13; Gaps 7;

QY 6 IVAEICNGSVVDIAREMILKAKAGVAVKQTFKADKLISALAPKAEYQIKTGELE 65
 DB 19 VVEEIGNHGSLBLAKIMVDAFSTGAKIKHQTHIVEDMSKAKV---IPGNAKI- 74
 QY 66 SOLEMTKLEMKYDYLHMEYAVSLNLDVSTPFDESDIDFLASIKOKIMKIPSGELN 125
 DB 75 SIYEIMOKCALDYDELAKYTERKGLVYSTPFSRAGARLDEMGSAFKIGSGECNN 134
 QY 126 LPYLEKIAKLPIPDKKIISTGMATIDEIKOSVSIFINNKVPVGNITTLACNTEPTPE 185
 DB 135 YPLIKHIAF---KKPMIVSTGMNISIEIKPTVKILDNELP---FVLMHTTNLYPTPHN 188
 QY 186 DVNLAINDLKKEPPKNNIGFSDHSGFYALAAVPGYITFLKHTTLDKSMGPHLAS 245
 DB 189 LVRLNMLELKEF--SCWGLSDHTTDNLACGVALGACVLERHFTDSMRSGPDVCS 247
 QY 246 IEPDELKHLCTIGRCVCEKSLGNS-SKVVTASERKNKIYAKSITIAKTEIKGEVPEKKNI 304
 DB 248 MDQALKEKLTIOSEGMALMRGNESKKAQOEVTIDFAFASVVSIKDIKGEVLSMDNI 307
 QY 305 TTKRPG-NGISPMWYNLIGKIAEODFIPDELIIHSEF 341
 DB 308 WKRPGLGIGISAEPENILGKALRDIENDTQLSYEDF 345

RESULT 9
 US-10-303-128-35
 Sequence 35, Application US/10303128
 Publication No. US20030157656A1
 GENERAL INFORMATION:
 APPLICANT: Gilbert, Michel
 APPLICANT: Wakarchuk, Warren W.
 APPLICANT: National Research Council of Canada
 TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
 TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
 FILE REFERENCE: 019633-00011US
 CURRENT APPLICATION NUMBER: US/10/303,128
 PRIOR FILING DATE: 2002-11-21
 PRIOR APPLICATION NUMBER: US/09/816,028
 PRIOR FILING DATE: 2001-03-21
 PRIOR APPLICATION NUMBER: US 60/118,213
 PRIOR FILING DATE: 1999-02-01
 PRIOR APPLICATION NUMBER: US 09/495,406
 PRIOR FILING DATE: 2000-01-31
 NUMBER OF SEQ ID NOS: 49
 SOFTWARE: PatentIn Ver. 2.1
 SEQ ID NO 35
 LENGTH: 346
 TYPE: PRF
 ORGANISM: Campylobacter jejuni
 FEATURE:
 OTHER INFORMATION: stalic acid synthase from C. jejuni OH4384 (ORF 8a
 OTHER INFORMATION: of lipooligosaccharide (LOS) biosynthesis locus)
 US-10-303-128-35

Query Match 26.7%; Score 475.5; DB 14; Length 346;

Thu May 13 11:53:13 2004

us-09-930-440b-8.rapb

Page 5

Best Local Similarity 36.7%; Pred. No. 1.1e-36;
Matches 124; Conservative 54; Mismatches 147; Indels 13; Gaps 7;
QY 6 IVAIGCNHNSVDIAREMILKAEAGVNAVKQTFPAADKLISAIAPKAEYQIKNTGELE 65
DB 19 VVPEIGINNGSLAKIMVDAFSTGAKIIKHQTHIVEDMSKAAKV---IPGNAKI- 74
QY 66 SULEMTKLEKDYDYLHMEYAVSLNDVSTFPEDSDIDFLASLKOKIWKIPSGELIN 125
DB 75 SIYEIMOKCALDYDELAKYETKGLVYLSTFSPAGANRLEDMGVSAFKISGSCNN 134
QY 126 LPYIEKIAKLPIPDKIIISTGATIDEIKOSVIFINNKVPVGNITTLKNTETPTPE 185
DB 135 YPLIKHIAAF---KKPMIVSTGNSISIKPTVKILNDNEIP---FVIMHTTNLYPTPHN 188
QY 186 DVNINAINDLKKHPKKNIGFSDHSSGFYAAIAAVPYGITFEIKHFTLDSKMSGPDHLAS 245
DB 189 LVRLNMLLEKKEF-SCWVGLSDHTTNDLACGVALGACVLERHFTDSMERSGPDIVCS 247
QY 246 IEPDELKHLCTGVRCVKSIGSN-SKVVTASERKNKIIVAKSIITAKTEIKKGEVSEKNI 304
DB 248 MDTOALKEILLIIOSEQMAIMRGNNESKKAQEQVITIDFAPASVVISIKIKGEVLSMDNI 307
QY 305 TTKRPG-NGISPMWYNLLGKIAODPIPELIIHSEF 341
DB 308 WVKRPGIGISAAEFENILGKALRDIENDTOLSYEDF 345
RESULT 10
US-10-303-134-35
Sequence 35, Application US/10303134
Publication No. US20030157657A1
GENERAL INFORMATION:
APPLICANT: Gilbert, Michel
APPLICANT: Wakarchuk, Warren W.
TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
TITLE OF INVENTION: Gangliosides and Ganglioside Mimics
FILE REFERENCE: 019633-00011105
CURRENT APPLICATION NUMBER: US/10/303,134
CURRENT FILING DATE: 2002-11-21
PRIOR APPLICATION NUMBER: US/09/816,028
PRIOR FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: US 60/118,213
PRIOR FILING DATE: 1999-02-01
PRIOR APPLICATION NUMBER: US 09/495,406
PRIOR FILING DATE: 2000-01-31
NUMBER OF SEQ ID NOS: 49
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 35
LENGTH: 346
TYPE: PRT
ORGANISM: Campylobacter jejuni
FEATURE:
OTHER INFORMATION: sialic acid synthase from C. jejuni OH4384 (ORF 8a
OTHER INFORMATION: of lipooligosaccharide (LOS) biosynthesis locus)
US-10-303-134-35
Query Match 26.7%; Score 475.5; DB 14; Length 346;
Best Local Similarity 36.7%; Pred. No. 1.1e-36;
Matches 124; Conservative 54; Mismatches 147; Indels 13; Gaps 7;
QY 6 IVAIGCNHNSVDIAREMILKAEAGVNAVKQTFPAADKLISAIAPKAEYQIKNTGELE 65
DB 19 VVPEIGINNGSLAKIMVDAFSTGAKIIKHQTHIVEDMSKAAKV---IPGNAKI- 74
QY 66 SULEMTKLEKDYDYLHMEYAVSLNDVSTFPEDSDIDFLASLKOKIWKIPSGELIN 125
DB 75 SIYEIMOKCALDYDELAKYETKGLVYLSTFSPAGANRLEDMGVSAFKISGSCNN 134
QY 126 LPYIEKIAKLPIPDKIIISTGATIDEIKOSVIFINNKVPVGNITTLKNTETPTPE 185
DB 135 YPLIKHIAAF---KKPMIVSTGNSISIKPTVKILNDNEIP---FVIMHTTNLYPTPHN 188
QY 186 DVNINAINDLKKHPKKNIGFSDHSSGFYAAIAAVPYGITFEIKHFTLDSKMSGPDHLAS 245
DB 189 LVRLNMLLEKKEF-SCWVGLSDHTTNDLACGVALGACVLERHFTDSMERSGPDIVCS 247
QY 246 IEPDELKHLCTGVRCVKSIGSN-SKVVTASERKNKIIVAKSIITAKTEIKKGEVSEKNI 304
DB 248 MDTOALKEILLIIOSEQMAIMRGNNESKKAQEQVITIDFAPASVVISIKIKGEVLSMDNI 307
QY 305 TTKRPG-NGISPMWYNLLGKIAODPIPELIIHSEF 341
DB 308 WVKRPGIGISAAEFENILGKALRDIENDTOLSYEDF 345
RESULT 11
US-10-303-162-35
Sequence 35, Application US/10303162
Publication No. US20030157658A1
GENERAL INFORMATION:
APPLICANT: Gilbert, Michel
APPLICANT: Wakarchuk, Warren W.
TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
TITLE OF INVENTION: Gangliosides and Ganglioside Mimics
FILE REFERENCE: 019633-00011105
CURRENT APPLICATION NUMBER: US/10/303,162
CURRENT FILING DATE: 2002-11-21
PRIOR APPLICATION NUMBER: US/09/816,028
PRIOR FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: US 60/118,213
PRIOR FILING DATE: 1999-02-01
PRIOR APPLICATION NUMBER: US 09/495,406
PRIOR FILING DATE: 2000-01-31
NUMBER OF SEQ ID NOS: 49
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 35
LENGTH: 346
TYPE: PRT
ORGANISM: Campylobacter jejuni
FEATURE:
OTHER INFORMATION: sialic acid synthase from C. jejuni OH4384 (ORF 8a
OTHER INFORMATION: of lipooligosaccharide (LOS) biosynthesis locus)
US-10-303-162-35
Query Match 26.7%; Score 475.5; DB 14; Length 346;
Best Local Similarity 36.7%; Pred. No. 1.1e-36;
Matches 124; Conservative 54; Mismatches 147; Indels 13; Gaps 7;
QY 6 IVAIGCNHNSVDIAREMILKAEAGVNAVKQTFPAADKLISAIAPKAEYQIKNTGELE 65
DB 19 VVPEIGINNGSLAKIMVDAFSTGAKIIKHQTHIVEDMSKAAKV---IPGNAKI- 74
QY 66 SULEMTKLEKDYDYLHMEYAVSLNDVSTFPEDSDIDFLASLKOKIWKIPSGELIN 125
DB 75 SIYEIMOKCALDYDELAKYETKGLVYLSTFSPAGANRLEDMGVSAFKISGSCNN 134
QY 126 LPYIEKIAKLPIPDKIIISTGATIDEIKOSVIFINNKVPVGNITTLKNTETPTPE 185
DB 135 YPLIKHIAAF---KKPMIVSTGNSISIKPTVKILNDNEIP---FVIMHTTNLYPTPHN 188
QY 186 DVNINAINDLKKHPKKNIGFSDHSSGFYAAIAAVPYGITFEIKHFTLDSKMSGPDHLAS 245
DB 189 LVRLNMLLEKKEF-SCWVGLSDHTTNDLACGVALGACVLERHFTDSMERSGPDIVCS 247
QY 246 IEPDELKHLCTGVRCVKSIGSN-SKVVTASERKNKIIVAKSIITAKTEIKKGEVSEKNI 304
DB 248 MDTOALKEILLIIOSEQMAIMRGNNESKKAQEQVITIDFAPASVVISIKIKGEVLSMDNI 307
QY 305 TTKRPG-NGISPMWYNLLGKIAODPIPELIIHSEF 341
DB 308 WVKRPGIGISAAEFENILGKALRDIENDTOLSYEDF 345

QY 186 DVNINAINDLKKHPKKNIGFSDHSSGFYAAIAAVPYGITFEIKHFTLDSKMSGPDHLAS 245
DB 189 LVRLNMLLEKKEF-SCWVGLSDHTTNDLACGVALGACVLERHFTDSMERSGPDIVCS 247
QY 246 IEPDELKHLCTGVRCVKSIGSN-SKVVTASERKNKIIVAKSIITAKTEIKKGEVSEKNI 304
DB 248 MDTOALKEILLIIOSEQMAIMRGNNESKKAQEQVITIDFAPASVVISIKIKGEVLSMDNI 307
QY 305 TTKRPG-NGISPMWYNLLGKIAODPIPELIIHSEF 341
DB 308 WVKRPGIGISAAEFENILGKALRDIENDTOLSYEDF 345
RESULT 11
US-10-303-162-35
Sequence 35, Application US/10303162
Publication No. US20030157658A1
GENERAL INFORMATION:
APPLICANT: Gilbert, Michel
APPLICANT: Wakarchuk, Warren W.
TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
TITLE OF INVENTION: Gangliosides and Ganglioside Mimics
FILE REFERENCE: 019633-00011105
CURRENT APPLICATION NUMBER: US/10/303,162
CURRENT FILING DATE: 2002-11-21
PRIOR APPLICATION NUMBER: US/09/816,028
PRIOR FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: US 60/118,213
PRIOR FILING DATE: 1999-02-01
PRIOR APPLICATION NUMBER: US 09/495,406
PRIOR FILING DATE: 2000-01-31
NUMBER OF SEQ ID NOS: 49
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 35
LENGTH: 346
TYPE: PRT
ORGANISM: Campylobacter jejuni
FEATURE:
OTHER INFORMATION: sialic acid synthase from C. jejuni OH4384 (ORF 8a
OTHER INFORMATION: of lipooligosaccharide (LOS) biosynthesis locus)
US-10-303-162-35
Query Match 26.7%; Score 475.5; DB 14; Length 346;
Best Local Similarity 36.7%; Pred. No. 1.1e-36;
Matches 124; Conservative 54; Mismatches 147; Indels 13; Gaps 7;
QY 6 IVAIGCNHNSVDIAREMILKAEAGVNAVKQTFPAADKLISAIAPKAEYQIKNTGELE 65
DB 19 VVPEIGINNGSLAKIMVDAFSTGAKIIKHQTHIVEDMSKAAKV---IPGNAKI- 74
QY 66 SULEMTKLEKDYDYLHMEYAVSLNDVSTFPEDSDIDFLASLKOKIWKIPSGELIN 125
DB 75 SIYEIMOKCALDYDELAKYETKGLVYLSTFSPAGANRLEDMGVSAFKISGSCNN 134
QY 126 LPYIEKIAKLPIPDKIIISTGATIDEIKOSVIFINNKVPVGNITTLKNTETPTPE 185
DB 135 YPLIKHIAAF---KKPMIVSTGNSISIKPTVKILNDNEIP---FVIMHTTNLYPTPHN 188
QY 186 DVNINAINDLKKHPKKNIGFSDHSSGFYAAIAAVPYGITFEIKHFTLDSKMSGPDHLAS 245
DB 189 LVRLNMLLEKKEF-SCWVGLSDHTTNDLACGVALGACVLERHFTDSMERSGPDIVCS 247
QY 246 IEPDELKHLCTGVRCVKSIGSN-SKVVTASERKNKIIVAKSIITAKTEIKKGEVSEKNI 304
DB 248 MDTOALKEILLIIOSEQMAIMRGNNESKKAQEQVITIDFAPASVVISIKIKGEVLSMDNI 307
QY 305 TTKRPG-NGISPMWYNLLGKIAODPIPELIIHSEF 341
DB 308 WVKRPGIGISAAEFENILGKALRDIENDTOLSYEDF 345

RESULT 12
US-10-156-761-10910
Sequence 10910, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBAH, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 10910
LENGTH: 311
TYPE: PRT
ORGANISM: Streptomyces avermitilis
US-10-156-761-10910
Query Match 24.0%; Score 426; DB 14; Length 311;
Best Local Similarity 33.9%; Pred. No. 5.1e-22;
Matches 100; Conservative 60; Mismatches 117; Indels 18; Gaps 6;
DB
QY 4 IYVAEIGCHNGSVDIAREMILKAEAGVAVKQFTFKADKLISAPRAEYOIKNT--61
20 VYVVGELGINHNGELKNAFLIDAAAGCAVAKFQ---KRTPEICTPDQMDIERDTP 75
DB 62 -GELSQLEMTKLEMYDYLLMEYAVSLNLDVSTPDESDISFLSKQIMKIPS 120
76 WGRW-TIIDYRHRREFEEDYRQIDETAKSKNDWPSPDITAAVAFLEKEDIPAKKVAS 134
QY 121 GELMLPYLEKIAKLPIPDKKIISTGMATIDELIKOSVSIFNNKVPVGNITILHCNTEY 180
135 ASLTLD--DELLRALRGTCRTVILSTGMSYKQIRHAEVETASD---NILLCHASTVY 186
DB 181 PTPEDVNAINDLKHPKNNIGPSDSSGFFAAIAVPYGTTFLEKFTLDKSMGDP 240
187 PAKAEELNLKVINLLOEYFNVPVIGSGHETGLQTLIAVAALANTFVERHITIDRAMNGS 246
QY 241 DHAASIEPDELKHLCTGRCVEKSLGNSKVTASE--RNKKIYARKSIIIAKTEI 293
247 DQAAVPEPQGLTRLVDRIDRTTERSLGDKVYKVESELGPMKKILRPVGVAAEARI 301
DB
RESULT 13
US-09-882-227-490
Sequence 490, Application US/09882227
Publication No. US20030158396A1
GENERAL INFORMATION:
APPLICANT: Kleantous, Harold
APPLICANT: Al-Garawi, Amal
APPLICANT: Miller, Charles
APPLICANT: Tomb, Jean-Francois
APPLICANT: Ooomen, Raymond P.
TITLE OF INVENTION: Identification of Polynucleotides
TITLE OF INVENTION: Encoding No. US20030158396a1 Helicobacter Polypeptides in the
TITLE OF INVENTION: Genome
FILE REFERENCE: 06132/047002
CURRENT APPLICATION NUMBER: US/09/882,227
CURRENT FILING DATE: 2001-06-15
PRIOR APPLICATION NUMBER: US 08/902,615
PRIOR FILING DATE: 1997-07-29
NUMBER OF SEQ ID NOS: 638
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 490

LENGTH: 340
TYPE: PRT
ORGANISM: Helicobacter pylori
US-09-882-227-490
Query Match 22.6%; Score 402; DB 10; Length 340;
Best Local Similarity 31.8%; Pred. No. 1.1e-29;
Matches 104; Conservative 74; Mismatches 135; Indels 14; Gaps 8;
DB
QY 6 IVAEIGCHNGSVDIAREMILKAEAGVAVKQFTFKADKLISAPRAEYOIKNT--GEL 64
7 IYVAEISNHNQDNLAKESLHAIKESGADFYKQTY--TPSCCTLNKSKEDPFIQTLMDX 65
QY 65 ESQLEMTKLEMYDYLLMEYAVSLNLDVSTPDESDISFLSKQIMKIPSGLT 124
66 EMLYELYOKASTPEWMAELFELARKDGLIFSSPSSQALELLESINCPWKIASFEIV 125
DB 125 NLPEYKIAKLPIPDKKIISTGMATIDELIKOSVSIF--INNRYVGNITILHCNTEYPT 182
126 DIDLIEKAR--YKPIIILSSGIATHELQDAISLCRRVNN---PITILKCVSAYPS 178
QY 183 PTPEDVNAINDLKHPKNNIGPSDSSGFFAAIAVPYGTTFLEKFTLDKSMGDPH 242
179 KIEDANLISMYVLGEIFGV--KFGLSDHITIGSLCPILATTLGASMEKHFILMKSLQTPDS 237
DB 243 LASIEPDELKHLCTGRCVEKSLGNS--KVTASERKKIYARKSIIIAKTEIKGSEVSE 301
238 AFSDMFGKSWVEAIKOSVTLGSEEPRIINPKTEKRFRFASLFLVK-DIQGEALTE 296
QY 302 KNITTRPGNGISPMENYLLGKIAEQ 328
DB 297 NIKKALRPVGLHPKFEKELIGQKASK 323
DB
RESULT 14
US-10-335-977-7626
Sequence 7626, Application US/10335977
Publication No. US20040052799A1
GENERAL INFORMATION:
APPLICANT: DOUGLAS SMITH et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
RELATING TO HELICOBLACTER PYLORI FOR
DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 10031
CORRESPONDENCE ADDRESS:
ADDRESSER: LAHIVE & COCKFIELD
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: Windows NT 4.0
SOFTWARE: UNIX
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/335,977
FILING DATE: 30-Dec-2002
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/993,002
FILING DATE: 17-DEC-1997
ATTORNEY/AGENT INFORMATION:
NAME: Mandragoras, Amy E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: GTN-018
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)742-4214
INFORMATION FOR SEQ ID NO: 7626:
SEQUENCE CHARACTERISTICS:
LENGTH: 340 amino acids
TYPE: amino acid

Thu May 13 11:53:13 2004

us-09-930-440b-8.rapb

Page 7

TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Helicobacter pylori
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...340
SEQUENCE DESCRIPTION: SEQ ID NO: 7626
US-10-335-977-7626

Query Match 21.4%; Score 381; DB 12; Length 340;
Best Local Similarity 30.1%; Pred. No. 1,2e-27;
Matches 98; Conservative 75; Mismatches 141; Indels 12; Gaps 6;

QY 6 IYAEIGCNHNSVDIAREMILKAKAGVNAVVFOTFKADKLISALAPKAEYQIKNT-GEL 64
DB 7 IYAEISANHQDINLAKESLHAIKESGADPVKLQY-TPSQMTLDSKEDPFIIOGTLMK 65
QY 65 ESQLEMTKLEMKYDDYLHMEYAVSLNDVSTPDEDSDIDFLASLKQIKIPSGELL 124
DB 66 ENLYGYOKRSTPLEMHAELFELAKKLDLGISSPSSKALELSDCEMTKIASFEIV 125
QY 125 NLYLEKIKALPIPDKKIISTGMATIDEIKOSVSIF--INNRPVGNITILHCTEYPT 182
DB 126 DDLLEKAR--TQKPIILSSGIAHAELODAISLCRGVNN---FDITLLKVSAYPS 178
QY 183 PREDVNAININDLKKEPPKNNIGFSQHSQFYAALAAVPGITFTIEKHTLDSKSGPDH 242
DB 179 KIDDAHLASVVKLGTEFGV-KFGISDHTIGSLCPILATTLGASVTEHFLINKSLQTPDS 237
QY 243 LASIPEDELKHLCTIGVRCVSKLSGNSKRYVTASERKKIVAKRSIIAKTEIRKGEVSEK 302
DB 238 AFSMDPNGFSMWGALKOSVIALGEEEPKINPKTLERRFPARSLVVIDIOGGEALTSD 297
QY 303 NITTKRPGNGISPMENYNLGKIAEQ 328
DB 298 NIKALRPNLGHPKFEYKEIIGOKASK 323

RESULT 15
US-10-335-977-7625
Sequence 7625, Application US/10335977
Publication No. US20040052799A1
GENERAL INFORMATION:
APPLICANT: DOUGLAS SMITH et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
RELATING TO HELICOBLACTER PYLORI FOR
DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 10031
CORRESPONDENCE ADDRESS:
ADDRESS: LAHIVE & COCKFIELD
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: Windows NT 4.0
SOFTWARE: UNIX
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/335,977
FILING DATE: 30-Dec-2002
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/993,002
FILING DATE: 17-Dec-1997
ATTORNEY/AGENT INFORMATION:
NAME: Mandagoudar, Amy E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: GTN-018
TELECOMMUNICATION INFORMATION:

TELEPHONE: (617)227-7400
TELEFAX: (617)742-4214
INFORMATION FOR SEQ ID NO: 7625:
SEQUENCE CHARACTERISTICS:
LENGTH: 314 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Helicobacter pylori
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...314
SEQUENCE DESCRIPTION: SEQ ID NO: 7625:
US-10-335-977-7625

Query Match 21.0%; Score 373; DB 12; Length 314;
Best Local Similarity 30.4%; Pred. No. 6e-27;
Matches 95; Conservative 72; Mismatches 134; Indels 12; Gaps 6;

QY 6 IYAEIGCNHNSVDIAREMILKAKAGVNAVVFOTFKADKLISALAPKAEYQIKNT-GEL 64
DB 11 IYAEISANHQDINLAKESLHAIKESGADPVKLQY-TPSQMTLDSKEDPFIIOGTLMK 69
QY 65 ESQLEMTKLEMKYDDYLHMEYAVSLNDVSTPDEDSDIDFLASLKQIKIPSGELL 124
DB 70 ENLYGYOKRSTPLEMHAELFELAKKLDLGISSPSSKALELSDCEMTKIASFEIV 129
QY 125 NLYLEKIKALPIPDKKIISTGMATIDEIKOSVSIF--INNRPVGNITILHCTEYPT 182
DB 130 DDLLEKAR--TQKPIILSSGIAHAELODAISLCRGVNN---FDITLLKVSAYPS 182
QY 183 PREDVNAININDLKKEPPKNNIGFSQHSQFYAALAAVPGITFTIEKHTLDSKSGPDH 242
DB 183 KIDDAHLASVVKLGTEFGV-KFGISDHTIGSLCPILATTLGASVTEHFLINKSLQTPDS 241
QY 243 LASIPEDELKHLCTIGVRCVSKLSGNSKRYVTASERKKIVAKRSIIAKTEIRKGEVSEK 302
DB 242 AFSMDPNGFSMWGALKOSVIALGEEEPKINPKTLERRFPARSLVVIDIOGGEALTSD 301
QY 303 NITTKRPGNGISP 315
DB 302 NIKALRPNLGVT 314

Search completed: May 6, 2004, 09:25:08
Job time: 38.4054 secs